

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 6, 2005, 06:52:09 ; Search time 1311 Seconds
(without alignments)
11162.167 Million cell updates/sec

Title: US-10-616-263-29
Perfect score: 2472
Sequence: 1 tatgagccttcggaacttgt.....aaaaaaaaaaaaaaaaaaaaa 2472

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4390206 segs, 2959870667 residues
Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N Geneseq_16Dec04:*
1: Geneseqn1980s:*
2: Geneseqn1990s:*
3: Geneseqn2000s:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002as:*
7: Geneseqn2002bs:*
8: Geneseqn2003as:*
9: Geneseqn2003bs:*
10: Geneseqn2003cs:*
11: Geneseqn2003ds:*
12: Geneseqn2004as:*
13: Geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2472	100.0	2472	3	AA16632 Human sec
2	2467.2	99.8	2472	6	ABK35878 cDNA sequ
3	2406.6	97.4	2572	6	ABQ54852 Human ova
4	2399.2	97.1	2776	3	AA08567 Human cha
5	1961.6	79.4	2562	5	AA83823 DNA encod
6	1763.2	71.3	1921	3	AA09959 Human sec
7	1503.8	60.8	1536	3	AA077414 Human ORF
8	1314.2	53.2	1546	13	ADP43721 Human PMM
9	1253.8	50.7	1274	6	AA099977 Human sec
10	593	24.0	694	6	ABK35896 cDNA sequ
11	422.4	17.1	458	2	AA097580 Extended
12	422.4	17.1	458	12	ADP18847 Human sec
13	413.4	16.7	421	3	AA077867 cDNA enco
14	413.4	16.7	421	4	AA128605 Colon tum
15	413.4	16.7	421	8	AB232791 Human col
16	352	14.2	619	10	ADD71001 Human Mck
17	344.6	13.9	348	3	AA02899 Human sec
18	327.8	13.3	371	2	AA086645 EST clone
19	269	10.9	332	4	AA058781 cDNA #145
20	254.8	10.3	350	2	AA026699 Human gen

21	244.2	9.9	268	3	AAA41503
22	186.4	7.5	313	2	AA041298 Human sec
23	186	7.5	233	3	AA045883 Human sec
24	130.4	5.3	248	2	AA041081 Human sec
25	123	5.0	134	2	AA085281 Human sin
26	123	5.0	134	2	AA085282 Human sin
27	113.4	4.6	407	13	ADQ57549 Novel can
28	95.6	3.9	385	5	ADL37765 Human ova
29	95.6	3.9	385	5	ADL37765 Human ova
30	95.4	3.9	2136	11	ACN92757 Breast ca
31	94.6	3.8	534	13	ACN58719 Cotton gy
32	93.6	3.8	546	13	ADL62275 Cotton cd
33	93	3.8	349	5	ADL44149 Human ova
34	93	3.8	483	5	ABV59155 Human pro
35	93	3.8	1767	12	ADQ62272 Transcrip
36	92.4	3.7	469	13	ACN62049 Cotton gy
37	91.8	3.7	381	13	ACN55708 Cotton an
38	91.8	3.7	487	13	ACN48060 Cotton pr
39	91.6	3.7	3420	12	ADH61306 INTSIG en
40	91.4	3.7	390	5	ADL43918 Human ova
41	91.4	3.7	1762	6	AA145656 Human can
42	91.4	3.7	1762	6	AA145657 Human can
43	91.2	3.7	318	13	ACN51559 Cotton an
44	91.2	3.7	549	13	ACN52320 Cotton an
45	91	3.7	510	13	ADR64316 Cotton cd

ALIGNMENTS

RESULT 1

AAA16632
ID AAA16632 standard; cDNA; 2472 BP.
AC AAA16632;
XX
XX 16-JUN-2000 (first entry)
XX Human secreted protein clone pp314_19 nucleotide sequence SEQ ID NO:29.
DE Human; secreted protein; immunostimulant; immunosuppressant; virucide;
XX antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;
KW antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;
KW antithyroid; immune deficiency; severe combined immunodeficiency; SCID;
KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
KW connective tissue disease; multiple sclerosis; erythematosis;
KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
KW insulin dependent diabetes mellitus; graft-versus-host-disease;
KW autoimmune inflammatory eye disease; allergy; ss.
XX Homo sapiens.
OS
XX WO200009552-A1.
PN
XX 24-FEB-2000.
PD
XX 13-AUG-1999; 99WO-US018298.
PF
XX 14-AUG-1998; 98US-0096622P.
PR 17-AUG-1998; 98US-0096815P.
PR 04-SEP-1998; 98US-0099229P.
PR 23-OCT-1998; 98US-0105368P.
PR 08-JAN-1999; 99US-0115234P.
PR 12-FEB-1999; 99US-0119931P.
PR 18-FEB-1999; 99US-0120575P.
PR 30-APR-1999; 99US-0132020P.
PR 11-AUG-1999; 99US-0148424P.
XX (GENY) GENETICS INST INC.
PA
XX Jacobs K, McCoey JM, Lavallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;

PI Wong GG, Clark HF, Fechtel K;
XX WPI: 2000-205979/18.
DR P-PSDB; AAY94912.
XX
PT New polynucleotides encoding secreted proteins, which may have e.g.
PT nutritional, chemokine, immune stimulating or suppressing, hematopoiesis
PT regulating, tissue growth, activin/inhibin antiinflammatory or tumor
PT inhibition activity.
XX
PS Claim 38; Page 498-499; 641bp; English.
XX
XX AAA16618 to AAA16697 encode the human secreted proteins given in AAY94898
CC to AAY94980, isolated from human adult brain, adult thyroid, adult
CC retina, foetal carcinoma, adult blood, adult neural, foetal kidney, adult
CC placenta, adult testis, whole embryo, adult cartilage, kidney, foetal
CC brain, adult thymus, foetal placenta, adult uterus, adult tumour, and
CC adult bladder, cDNA libraries. The polynucleotides and proteins are
CC predicted to have biological activities which would make them suitable
CC for treating, preventing or ameliorating medical conditions in humans and
CC animals. The polynucleotides can be used as markers for tissues in which
CC the protein is preferentially expressed, as molecular weight markers on
CC Southern gels, and as chromosome markers or tags to identify chromosomes
CC or to map gene positions. The proteins can be used in the treatment of
CC immune deficiencies and disorders, such as severe combined
CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other
CC infections. These infections include human immunodeficiency virus (HIV),
CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
CC candidiasis. The proteins can be used to treat autoimmune disorders such
CC as connective tissue disease, multiple sclerosis, systemic lupus
CC erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation,
CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent
CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
CC autoimmune inflammatory eye disease. The proteins can also be used to
CC treat allergic conditions, such as asthma. AAA16698 to AAA16774 represent
CC probes for the human secreted proteins from the present invention
XX
SQ Sequence 2472 BP; 796 A; 475 C; 502 G; 699 T; 0 U; 0 Other;

Query Match 100.0%; Score 2472; DB 3; Length 2472;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2472; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TATGAGCCTTCGAACTTGGAGAGACTACAAAGTTTGGTTGTTATGGTCCCTTAGT 60
DB 1 TATGAGCCTTCGAACTTGGAGAGACTACAAAGTTTGGTTGTTATGGTCCCTTAGT 60

QY 61 TGGGCTCATACATTTGGGGTGGTACAGAAATCAAAAGCAGCCCTGTGTTCCAAATACCTAA 120
DB 61 TGGGCTCATACATTTGGGGTGGTACAGAAATCAAAAGCAGCCCTGTGTTCCAAATACCTAA 120

QY 121 AAACGACGACATTCCTGAGCAAGATAGTCTGGGACTTTCAAATCTTCAGAGAGCCAAAT 180
DB 121 AAACGACGACATTCCTGAGCAAGATAGTCTGGGACTTTCAAATCTTCAGAGAGCCAAAT 180

QY 181 CAGGGGGAAGTACAGCCTTGCAATCTTCAGGTAAAGAGCAGCTTTGAAATCTGAGCTTC 240
DB 181 CAGGGGGAAGTACAGCCTTGCAATCTTCAGGTAAAGAGCAGCTTTGAAATCTGAGCTTC 240

QY 241 ATATCGAAAGACAGATGAAATATACAGTTGGATTAGAAAGAACTGCGCTTCCTTAGCT 300
DB 241 ATATCGAAAGACAGATGAAATATACAGTTGGATTAGAAAGAACTGCGCTTCCTTAGCT 300

QY 301 GGGATATCTTTATAGGTGTCATCTTGGAAACATACCTTTTTCAGAGGTCGCAAGAGCAG 360
DB 301 GGGATATCTTTATAGGTGTCATCTTGGAAACATACCTTTTTCAGAGGTCGCAAGAGCAG 360

QY 361 TCTGTAATATTTTCAGTCTCAAGCAACAAAGAGATTTGAAGAGTGAAGTAAATAAT 420
DB 361 TCTGTAATATTTTCAGTCTCAAGCAACAAAGAGATTTGAAGAGTGAAGTAAATAAT 420

QY 421 ATTTGGAATTTACTAATTTGTCAATTAATCTATGCTGATTAGCTTCATAAATTCGA 480
DB 421 ATTTGGAATTTACTAATTTGTCAATTAATCTATGCTGATTAGCTTCATAAATTCGA 480

DB 421 ATTTGGAATTTACTAATTTGTCAATTAATCTATGCTGATTAGCTTCATAAATTCGA 480
QY 481 ACTTTTGTGATTTTATAGCCACAATGCTGCATATTTCACTTTAAATTCCTTAAAGATAAT 540
DB 481 ACTTTTGTGATTTTATAGCCACAATGCTGCATATTTCACTTTAAATTCCTTAAAGATAAT 540
QY 541 TTTTAATGTTAAACCGTGATAATGCAATAAATAGAAAATGTGGTTTACAAAATAAAAACG 600
DB 541 TTTTAATGTTAAACCGTGATAATGCAATAAATAGAAAATGTGGTTTACAAAATAAAAACG 600
QY 601 GTCTTCACCTAGTTACCACTGAAGTAAAGATGTCTGTTTGGAGCTGAAGAGCCATCAT 660
DB 601 GTCTTCACCTAGTTACCACTGAAGTAAAGATGTCTGTTTGGAGCTGAAGAGCCATCAT 660
QY 661 GTGTAGAGTGAACCACTGACAACTGAGAGAGTCAAGGACCACTTTTGTCTTGAAGAG 720
DB 661 GTGTAGAGTGAACCACTGACAACTGAGAGAGTCAAGGACCACTTTTGTCTTGAAGAG 720
QY 721 AATTGTACATCATGCTATGCGCCCTCAGGTAGGCTGAGCAGCTGCACAATGGCTTGG 780
DB 721 AATTGTAAACATCATGCTATGCGCCCTCAGGTAGGCTGAGCAGCTGCACAATGGCTTGG 780
QY 781 AGGTTACGTTGTACAACTCAGTCTCAGCTCTGCTCAGTCACTTTTGGTGCACACA 840
DB 781 AGGTTACGTTGTACAACTCAGTCTCAGCTCTGCTCAGTCACTTTTGGTGCACACA 840
QY 841 TCCCATTTTAAAGATCCTGACAGCCTCCATAAGAAATCATGTGTCAAGCTTCAGTGAT 900
DB 841 TCCCATTTTAAAGATCCTGACAGCCTCCATAAGAAATCATGTGTCAAGCTTCAGTGAT 900
QY 901 TGGCTTATTTACAGCTATTTCTTGTGCTGCACTGAAATGTTGAAGATTTAGGCTT 960
DB 901 TGGCTTATTTACAGCTATTTCTTGTGCTGCACTGAAATGTTGAAGATTTAGGCTT 960
QY 961 GACACCCACCATGCTAGTAAATTAATAAATCAATCTTTTGGATCTTTTGCATCAGTTAT 1020
DB 961 GACACCCACCATGCTAGTAAATTAATAAATCAATCTTTTGGATCTTTTGCATCAGTTAT 1020
QY 1021 CAAAGTCTGATACCTGTGGTTGCAATCCAGTGGACTTTTAGTAGTACTCAGATCCTCCT 1080
DB 1021 CAAAGTCTGATACCTGTGGTTGCAATCCAGTGGACTTTTAGTAGTACTCAGATCCTCCT 1080
QY 1081 TTGTTTGGTGGCTAGTATTAACAAAGTAAACCTGCTGTATGCTCACCAGAAAGGAAC 1140
DB 1081 TTGTTTGGTGGCTAGTATTAACAAAGTAAACCTGCTGTATGCTCACCAGAAAGGAAC 1140
QY 1141 AGAGCATGTGCTGCTTTGATCCTCAGAGCCTTTTGTCTTACAAATTCAGAAAAATGCTGA 1200
DB 1141 AGAGCATGTGCTGCTTTGATCCTCAGAGCCTTTTGTCTTACAAATTCAGAAAAATGCTGA 1200
QY 1201 AGGCCACATCATTTTAGGAAAGATTTAATTTGTAACCTTTTAAAGGTCAGAAAGATTAGA 1260
DB 1201 AGGCCACATCATTTTAGGAAAGATTTAATTTGTAACCTTTTAAAGGTCAGAAAGATTAGA 1260
QY 1261 TTCCACTGTATTTACCTGGGATACTCAATGAAATGTCAGAAATCAATTAATGAGGCTATT 1320
DB 1261 TTCCACTGTATTTACCTGGGATACTCAATGAAATGTCAGAAATCAATTAATGAGGCTATT 1320
QY 1321 ACTATCAAAAAATCAACTGCCCTCAAGGTGGCACTCTTTTGTACAACCTTTTATCCGGAGA 1380
DB 1321 ACTATCAAAAAATCAACTGCCCTCAAGGTGGCACTCTTTTGTACAACCTTTTATCCGGAGA 1380
QY 1381 CACTTCTGACACTGGAGAGGAACTGTGGTGTGCTGATTTATGGGTTTCTCTTGAAGATGC 1440
DB 1381 CACTTCTGACACTGGAGAGGAACTGTGGTGTGCTGATTTATGGGTTTCTCTTGAAGATGC 1440
QY 1441 AGCTTTGGACCAAGCTGCTTAACTAGGAGGAGCTTAATCAGTGCACCAAGTATCTTCT 1500
DB 1441 AGCTTTGGACCAAGCTGCTTAACTAGGAGGAGCTTAATCAGTGCACCAAGTATCTTCT 1500
QY 1501 CCGTGCCAAAAAGTTATACATCCATCTTTTGAAGCAGTTTCTCAATATGATCGATTTAT 1560
DB 1501 CCGTGCCAAAAAGTTATACATCCATCTTTTGAAGCAGTTTCTCAATATGATCGATTTAT 1560

QY 1561 TGCATAGACAGAAATGGAGTGAATCTGTGATGGAAACCCCTGACTAAATGACAGGAACACA 1620
DB |||||
QY 1561 TGCATAGACAGAAATGGAGTGAATCTGTGATGGAAACCCCTGACTAAATGACAGGAACACA 1620
DB |||||
QY 1621 GCCTATTGGATCCCTAGGCTCAATATGCTCTAATAGTTATGGAAGTGTGAAGATGTGTG 1680
DB |||||
QY 1621 GCCTATTGGATCCCTAGGCTCAATATGCTCTAATAGTTATGGAAGTGTGAAGATGTGTG 1680
DB |||||
QY 1681 CACTGCAAAATTTGGCTCCCAACAATTTTTCATCTTATTCCTAAATGAAGCAACAATCTG 1740
DB |||||
QY 1681 CACTGCAAAATTTGGCTCCCAACAATTTTTCATCTTATTCCTAAATGAAGCAACAATCTG 1740
DB |||||
QY 1741 CAGCTTGCTTCTCTGCAACAGAAATGACACTGCTGGGATGAGCTGAAGCTCAGGTGTC 1800
DB |||||
QY 1741 CAGCTTGCTTCTCTGCAACAGAAATGACACTGCTGGGATGAGCTGAAGCTCAGGTGTC 1800
DB |||||
QY 1801 GACGGCACTGCATGCTCTGCAAGTAACTCAAGCAACCATGGGCTTTGTTGGGAGGTGG 1860
DB |||||
QY 1801 GACGGCACTGCATGCTCTGCAAGTAACTCAAGCAACCATGGGCTTTGTTGGGAGGTGG 1860
DB |||||
QY 1861 CTGTACTGAAATCATTTTGGCTGCATATATCAGACACAGACTCACAAGACCCAGAAAG 1920
DB |||||
QY 1861 CTGTACTGAAATCATTTTGGCTGCATATATCAGACACAGACTCACAAGACCCAGAAAG 1920
DB |||||
QY 1921 CATTTCTCAAGATGATGAATGTACTCAAAACAGAACTTCAATTAATTGCTGAAGCAATTTTG 1980
DB |||||
QY 1921 CATTTCTCAAGATGATGAATGTACTCAAAACAGAACTTCAATTAATTGCTGAAGCAATTTTG 1980
DB |||||
QY 1981 CAGTGGCTTAGAATCTGTTGGCTCTTTAGAACATGATGAGAGTGAATTTCTCACTGA 2040
DB |||||
QY 1981 CAGTGGCTTAGAATCTGTTGGCTCTTTAGAACATGATGAGAGTGAATTTCTCACTGA 2040
DB |||||
QY 2041 CATGAAGATGAGACACTTTGGTTCAGTTCAAGCAGATTTCTCCCTGTGTTGCTAACTGGCC 2100
DB |||||
QY 2041 CATGAAGATGAGACACTTTGGTTCAGTTCAAGCAGATTTCTCCCTGTGTTGCTAACTGGCC 2100
DB |||||
QY 2101 AGATTTGCTTTTCAAGTGTGGCTGTGATTTATCAATAGCCAGGAAGAACTCAACTGGTC 2160
DB |||||
QY 2101 AGATTTGCTTTTCAAGTGTGGCTGTGATTTATCAATAGCCAGGAAGAACTCAACTGGTC 2160
DB |||||
QY 2161 TTTCTTAAGACACAGCTGCTCATTTGTGACAAAGCTGCTTCCATGAAAGCTGT 2220
DB |||||
QY 2161 TTTCTTAAGACACAGCTGCTCATTTGTGACAAAGCTGCTTCCATGAAAGCTGT 2220
DB |||||
QY 2221 GGGCTCAGCCAGCAACTGACCTTGGACTGTTTGTGACCAAGCTTGTGGCTCAGCT 2280
DB |||||
QY 2221 GGGCTCAGCCAGCAACTGACCTTGGACTGTTTGTGACCAAGCTTGTGGCTCAGCT 2280
DB |||||
QY 2281 GGGCTTAGAGACAGCCAAATTTGATTTGGGATCTTTTCATATGTTTGAAGATAAAACTA 2340
DB |||||
QY 2281 GGGCTTAGAGACAGCCAAATTTGATTTGGGATCTTTTCATATGTTTGAAGATAAAACTA 2340
DB |||||
QY 2341 AGAAGATAGCATGTTGCTATTATCAAGAGAAACAAATTAATAGTCTGTGTTGGCAATTTGAGA 2400
DB |||||
QY 2341 AGAAGATAGCATGTTGCTATTATCAAGAGAAACAAATTAATAGTCTGTGTTGGCAATTTGAGA 2400
DB |||||
QY 2401 AA 2460
DB |||||
QY 2401 AA 2460
DB |||||
QY 2461 AAAAAAAAAAAAAA 2472
DB |||||
QY 2461 AAAAAAAAAAAAAA 2472

RESULT 2
ABK35878/c
ID ABK35878 standard; cDNA; 2472 BP.
XX
AC ABK35878;
XX
DT 08-MAY-2002 (first entry)

XX cDNA sequence #269 encoding novel human secreted protein.
DE Human secreted protein; hyperproliferative disorder; autoimmune disorder;
XX immune deficiency disorder; blood disorder; inflammatory disorder;
KW infectious disorder; allergic condition; neurodegenerative disorder;
KW liver fibrosis; coagulation disorder; gene therapy; antimicrobial;
KW tumour; cancer; hepatotropic; immunosuppressive; antirheumatic; gene; ss.
XX Homo sapiens.
OS
XX
XX WO200177289-A2.
XX
XX 18-OCT-2001.
XX
XX 29-MAR-2001; 2001WO-US010232.
XX
XX 06-APR-2000; 2000US-0195605P.
XX
XX (GEMY) GENETICS INST INC.
XX
XX Jacobs K, McCooy JM, Lavallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Agostino MJ, Bowman MR, Spaulding V, Wong GG;
PI Clark HF, Fechtel K, Howes SH, Resnick RJ, Gulukota K, Graham JR;
XX WPI; 2002-179322/23.
XX
XX Six hundred and twenty three polynucleotides derived from a variety of
PT human tissue sources which encode secreted proteins, useful for treating
PT immune deficiencies and disorders such as autoimmune disorders.
XX
XX Claim 1; Page 236-237; 393pp; English.
XX
XX The present invention relates to the isolation of novel cDNA sequences
CC which encode human secreted proteins. The cDNA sequences have been
CC derived from a variety of human tissues. The invention also provides a
CC method for producing proteins from these polynucleotide sequences. The
CC proteins are useful for identifying compounds that modulate their
CC activity and production. The sequences of the invention are useful for
CC treating diseases such as hyperproliferative disorders (e.g. cancer),
CC immune deficiency disorders (e.g. severe combined immunodeficiency
CC (SCID)), autoimmune disorders (e.g. multiple sclerosis), blood disorders
CC (e.g. thrombocytopenia), inflammatory disorders (e.g. arthritis),
CC infectious disorders (e.g. hepatitis), allergic conditions (e.g. asthma),
CC neurodegenerative disorders (e.g. Alzheimer's disease), liver fibrosis,
CC coagulation disorders (e.g. haemophilia), and tumours. The polynucleotide
CC sequences of the invention are also useful in gene therapy. ABK35610-
CC ABK36232 represent the cDNA sequences of the invention that encode for
CC novel human secreted proteins
XX
SQ Sequence 2472 BP; 699 A; 502 C; 474 G; 797 T; 0 U; 0 Other;

Query Match 99.8%; Score 2467.2; DB 6; Length 2472;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2469; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TATGAGCCCTCGGAACCTTGTGGAGAGACTACAAAGCTTTTGGTGTGATGGTCCCTTTAGT 60
DB |||||
DB 2472 TATGAGCCCTCGGAACCTTGTGGAGAGACTACAAAGCTTTTGGTGTGATGGTCCCTTTAGT 2413
QY 61 TGGGCTCATACATTTTGGGCTGGTACAGAAATCAAAAGCAGCCCTGTTTTCCAAATACCTAA 120
DB |||||
DB 2412 TGGGCTCATACATTTTGGGCTGGTACAGAAATCAAAAGCAGCCCTGTTTTCCAAATACCTAA 2353
QY 121 AAACGACGACATTCCTGAGCAAGATAGTCTGGGACTTTTCAAAATCTTCAGAGAGCCAAAT 180
DB |||||
DB 2352 AAACGACGACATTCCTGAGCAAGATAGTCTGGGACTTTTCAAAATCTTCAGAGAGCCAAAT 2293
QY 181 CCAGGGGAGTAGCAGGCTTGCAATCTTCAGGTAAAGAGCAGCTTTGATCTGAGCTTC 240
DB |||||
DB 2292 CCAGGGGAGTAGCAGGCTTGCAATCTTCAGGTAAAGAGCAGCTTTGATCTGAGCTTC 2233
QY 241 ATATCGAAAGAGAGATGAAAAATACCAAGTTGGATTAGAAAGAACTGGCTTCTTGTAGCT 300

Db 2232 ATATCGAAGAGGATGAAAATAACAGTTGGATTAGAAAGAACTTGGCTTCTTTAGCT 2173
QY 301 GGGATATCTTTTCATAGGTGTCATCTTTGGAAACATACCTTTTTCAGAGGTCTGCAAAAGCAG 360
Db 2172 GGGATATCTTTTCATAGGTGTCATCTTTGGAAACATACCTTTTTCAGAGGTCTGCAAAAGCAG 2113
QY 361 TCTGTAAATTTTCAGTCTCAAGCGAAACAAAGAGATTTGAAGAGTGAAGTAAATTAAT 420
Db 2112 TCTGTAAATTTTCAGTCTCAAGCGAAACAAAGAGATTTGAAGAGTGAAGTAAATTAAT 2053
QY 421 ATTTGGAAATTTACATTTTGTCAATTAATCATTTCTATGCTGATTAGCTTCATAAACATTGA 480
Db 2052 ATTTGGAAATTTACATTTTGTCAATTAATCATTTCTATGCTGATTAGCTTCATAAACATTGA 1993
QY 481 ACTTTTGTATTTTATAGCCCAATGCTGCATATTTTCATACCTTTAAATTTCTTAAAGAAATTAAT 540
Db 1992 ACTTTTGTATTTTATAGCCCAATGCTGCATATTTTCATACCTTTAAATTTCTTAAAGAAATTAAT 1933
QY 541 TTTTAATGTTTAAACGTTGATTAATGCAATAAATAGAAAATGTGGTTTACAAAATAAAAACG 600
Db 1932 TTTTAATGTTTAAACGTTGATTAATGCAATAAATAGAAAATGTGGTTTACAAAATAAAAACG 1873
QY 601 GTCTTCACTAGTTTACACCTGAAAGTAAAGATGTCTCGTTTGGAAAGCTAAGAGCCATCAT 660
Db 1872 GTCTTCACTAGTTTACACCTGAAAGTAAAGATGTCTCGTTTGGAAAGCTAAGAGCCATCAT 1813
QY 661 GTGTAAGAGTGAACCACTGACAACTGAGAGAGTCAGGACCAACATTTTCTGTGTAAGAAAG 720
Db 1812 GTGTAAGAGTGAACCACTGACAACTGAGAGAGTCAGGACCAACATTTTCTGTGTAAGAAAG 1753
QY 721 AATTGTAAACATCATGCTATGCGCCCTCAGTAGGCTGAAGCAGCTGCACAATGGCTTTGG 780
Db 1752 AATTGTAAACATCATGCTATGCGCCCTCAGTAGGCTGAAGCAGCTGCACAATGGCTTTGG 1693
QY 781 AGGTTACGTGTGTAACAACCTCAGCTCCTCAGCTCTGCTCAGTCACTTTTGGTTCACACA 840
Db 1692 AGGTTACGTGTGTAACAACCTCAGCTCCTCAGCTCTGCTCAGTCACTTTTGGTTCACACA 1633
QY 841 TCCCAATTTTAAAGATCCTGACAGCCCTCAATACAGAAATCATGTGTCAAGCTTCAGTGATTG 900
Db 1632 TCCCAATTTTAAAGATCCTGACAGCCCTCAATACAGAAATCATGTGTCAAGCTTCAGTGATTG 1573
QY 901 TGCGTTATTTACAGCTATTTCTTCTGCTGCAACCTGATTGAAATGTTTACAGATTAGGCTT 960
Db 1572 TGCGTTATTTACAGCTATTTCTTCTGCTGCAACCTGATTGAAATGTTTACAGATTAGGCTT 1513
QY 961 GACACCCACCACTGTCTATTAGATTAAATAAACATCTTTTGGAGTCTTTTGCATCAGTTATCT 1020
Db 1512 GACACCCACCACTGTCTATTAGATTAAATAAACATCTTTTGGAGTCTTTTGCATCAGTTATCT 1453
QY 1021 CAAGTCTGATACCTGTGGTTGTGCAATCCCAAGTGAATTTTGTAGTACTGAGATCTCTCCT 1080
Db 1452 CAAGTCTGAGACCTGTGGTTGTGCAATCCCAAGTGAATTTTGTAGTACTGAGATCTCTCCT 1393
QY 1081 TTCTTTGGTGGTGTATATTAACAAGTAAACCTGCTGTATGCTCACCAGAAAGAAAC 1140
Db 1392 TTCTTTGGTGGTGTATATTAACAAGTAAACCTGCTGTATGCTCACCAGAAAGAAAC 1333
QY 1141 AGAGCATGTCAAGTGTCTGAGAGCTTTTGTCTTACAAATTTCCAGAAAATGCTGA 1200
Db 1332 AGAGCATGTCAAGTGTCTGAGAGCTTTTGTCTTACAAATTTCCAGAAAATGCTGA 1273
QY 1201 AGGCCATCATTTTATAGGAAAGATTAAATTTGTATCCTTTTAAAGGTCAAAGATTATAGA 1260
Db 1272 AGGCCATCATTTTATAGGAAAGATTAAATTTGTATCCTTTTAAAGGTCAAAGATTATAGA 1213
QY 1261 TTCCATGTATTACCTGGGATCTCATTTGAAATGTGAGATTTCAATTAATGAGGCTATT 1320
Db 1212 TTCCATGTATTACCTGGGATCTCATTTGAAATGTGAGATTTCAATTAATGAGGCTATT 1153
QY 1321 ACCTATCAAAAAATCAACTGCCCTCAAGGTGGCACTCTTTTGTACAACTTTATCCCGAGA 1380

Db 1152 ACCTATCAAAAAATCAACTGCCCTCAAGGTGGCACTCTTTTGTACAACTTTATCCGAGA 1093
QY 1381 CACTTCTTGACACCTGGAGAAAGAACTGTGTGTGTGCTGATTATGGGGTTTCTCTTTGAAAAATGC 1440
Db 1092 CACTTCTTGACACCTGGAGAAAGAACTGTGTGTGTGCTGATTATGGGGTTTCTCTTTGAAAAATGC 1033
QY 1441 AGTCTTGGACCAAGCTGTAACTTAGGAGGAGCAGCTAACTCAGTGACCAAGTATAGTCTGT 1500
Db 1032 AGTCTTGGACCAAGCTGTAACTTAGGAGGAGCAGCTAACTCAGTGACCAAGTATAGTCTGT 973
QY 1501 CCGTGGCCAAAAAGTTATACATCCCATCTTTTGAAGCAGATTTCTCAATATGATCATCTGATTAT 1560
Db 972 CCGTGGCCAAAAAGTTATACATCCCATCTTTTGAAGCAGATTTCTCAATATGATCATCTGATTAT 913
QY 1561 TGCCATAGACAGAAATTTGGAGTGACTCTGTGATGAAACCCCTGACTAAAAATGACAGGAACACA 1620
Db 912 TGCCATAGACAGAAATTTGGAGTGACTCTGTATGAAACCCCTGACTAAAAATGACAGGAACACA 853
QY 1621 GCCTATTTGGATCCCTAGGCTCAATATGCTCAATAGTTATGGAAGTGTGAAAGATGTGTG 1680
Db 852 GCCTATTTGGATCCCTAGGCTCAATATGCTCAATAGTTATGGAAGTGTGAAAGATGTGTG 793
QY 1681 CACTGCAAAATTTGGCTCCCAACATTTTCTTCAATTTCTTAAATGAAGCAACAATCTG 1740
Db 792 CACTGCAAAATTTGGCTCCAAACATTTTCTTCAATTTCTTAAATGAAGCAACAATCTG 733
QY 1741 CAGCTTGTCTTCTGCAACAGAAATGACACTGCTGCTGGGATGAGCTGAAGCTCAGTGCTCA 1800
Db 732 CAGCTTGTCTTCTGCAACAGAAATGACACTGCTGCTGGGATGAGCTGAAGCTCAGTGCTCA 673
QY 1801 GACGCACTGCATGTCCTGCACTTAACTCAAGGAACCATGGCTTTTCTGGAGGTGG 1860
Db 672 GACGCACTGCATGTCCTGCACTTAACTCAAGGAACCATGGCTTTTCTGGAGGTGG 613
QY 1861 CTGTACTGAAATCTCATTTTGGCTGCATATATACAGACAAAGACTCAACAGCCAGCAAAAG 1920
Db 612 CTGTACTGAAATCTCATTTTGGCTGCATATATACAGACAAAGACTCAACAGCCAGCAAAAG 553
QY 1921 CATTTCTCAAAGATGATGATGATGATCTCAAACAGAACTTCAATTAATTTGCTGAAGCAATTTG 1980
Db 552 CATTTCTCAAAGATGATGATGATGATCTCAAACAGAACTTCAATTAATTTGCTGAAGCAATTTG 493
QY 1981 CAGTGCCTTAGAATCTGTGTGTGGCTCTTTAGAACTGATGAGGTGAATTTCTCACTCA 2040
Db 492 CAGTGCCTTAGAATCTGTGTGTGGCTCTTTAGAACTGATGAGGTGAATTTCTCACTCA 433
QY 2041 CATGAAGTATGACACCTTTTGTCTCAGTTTCAAGCAGATTTCTCCTGTGTGCTAACTGGCC 2100
Db 432 CATGAAGTATGACACCTTTTGTCTCAGTTTCAAGCAGATTTCTCCTGTGTGCTAACTGGCC 373
QY 2101 AGATTTGCTTTCACAGTGTGGCTGTGGATTATACAAATGACAGGAAGAACTCAACTGTGTC 2160
Db 372 AGATTTGCTTTCACAGTGTGGCTGTGGATTATACAAATGACAGGAAGAACTCAACTGTGTC 313
QY 2161 TTTTCTTAAAGACACACGTCGTCCTTTGTGCAACAAAGCTGCTTCCCATGTAAGCTGT 2220
Db 312 TTTTCTTAAAGACACACGTCGTCCTTTGTGCAACAAAGCTGCTTCCCATGTAAGCTGT 253
QY 2221 GGGCTCAGCAGCAACCTGACCTTTGAGCTGTTTGAAGTGAAGCTTGTAGTGGCTACAGGT 2280
Db 252 GGGCTCAGCAGCAACCTGACCTTTGAGCTGTTTGAAGTGAAGCTTGTAGTGGCTACAGGT 193
QY 2281 GGGCTCAGCAGCAACCTGACCTTTGAGCTGTTTGAAGTGAAGCTTGTAGTGGCTACAGGT 2340
Db 192 GGGCTCAGCAGCAACCTGACCTTTGAGCTGTTTGAAGTGAAGCTTGTAGTGGCTACAGGT 133
QY 2341 AGAGATAGCATGTTCTGATTACAAAGAGAAACAAATAAACTAGTCTGTGGCAATTTGACA 2400
Db 132 AGAGATAGCATGTTCTGATTACAAAGAGAAACAAATAAACTAGTCTGTGGCAATTTGACA 73
QY 2401 AAAAAAATAGCATGTTCTGATTACAAAGAGAAACAAATAAACTAGTCTGTGGCAATTTGACA 2460
Db 72 AAAAAAATAGCATGTTCTGATTACAAAGAGAAACAAATAAACTAGTCTGTGGCAATTTGACA 13

Qy 2461 AAAAAAAAAAAAA 2472:
 Db 12 AAAAAAAAAAAAA 1

RESULT 3
 ABQ54852
 ID ABQ54852 standard; cDNA; 2572 BP.
 XX
 AC ABQ54852;
 XX
 DT 22-AUG-2002 (first entry)
 XX
 DE Human ovarian antigen HMAEL73 cDNA, SEQ ID NO:732.
 XX
 KW Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;
 KW ovarian cancer; breast cancer; tumour; reproductive system disorder;
 KW infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;
 KW PCOS; ovarian cyst; dysmenorrhea; endocrine disorder; infection;
 KW inflammatory condition; immune disorder; blood disorder;
 KW cardiovascular disorder; respiratory disorder; neurological disorder;
 KW gastrointestinal disorder; urinary system disorder; drug screening;
 KW gene therapy; chromosome mapping; forensic analysis;
 KW antibody preparation; cytostatic; immunomodulatory; neuroprotective;
 KW antiinflammatory; gynaecological; reproductive; gene; ss.
 XX
 OS Homo sapiens.
 XX
 XX WO200200677-A1.
 PN
 XX
 PD 03-JAN-2002.
 XX
 PF 07-JUN-2001; 2001WO-US018569.
 XX
 PR 07-JUN-2000; 2000US-0209467P.
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA
 XX
 PI Birse CE, Rosen CA;
 XX
 DR WPI; 2002-147878/19.
 XX
 P-PSDB; ABP41775.
 XX
 XX Isolated nucleic acid molecules encoding novel ovarian polypeptides,
 PT useful in the prevention, treatment and diagnosis of cancer (e.g. ovarian
 PT cancer), immune disorders, cardiovascular disorders and neurological
 PT diseases.

Claim 1; SEQ ID NO 732; 2922pp; English.

The invention relates to 2175 novel human ovarian antigens (ABP41054-
 CC ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56303), and also
 CC encompasses polypeptides 90% identical and polynucleotides 95% identical
 CC to the sequences of the invention. The invention additionally relates to
 CC recombinant vectors and host cells comprising human ovarian antigen
 CC polynucleotides, antibodies against human ovarian antigens, and the use
 CC of ovarian antigen polynucleotides and polypeptides in diagnosing,
 CC treating, prognosing or preventing various ovary and/or breast-related
 CC disorders. Such conditions include ovarian cancer and breast cancer, and
 CC metastatic tumours of ovarian or breast origin, reproductive system
 CC disorders (e.g., infertility, disorders of pregnancy, anovulation,
 CC polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine
 CC disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic
 CC shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and
 CC vaginitis), immune disorders (e.g., congenital and acquired
 CC immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),
 CC blood-related disorders (e.g., anaemia), cardiovascular disorders,
 CC respiratory disorders, neurological disorders, gastrointestinal disorders
 CC and urinary system disorders. Ovarian antigen polypeptides and
 CC polynucleotides may also be used in screening for compounds which
 CC modulate ovarian antigen expression or activity. The polynucleotides may
 CC further be used for gene therapy, chromosome mapping, in the

CC identification of individuals and in forensic analysis, and the
 CC polypeptides may be used as food additives or to prepare antibodies
 CC useful in disease diagnosis, drug targeting and phenotyping. The present
 CC sequence represents cDNA encoding a human ovarian antigen of the
 CC invention. Note: the sequence data for this patent did not form part of
 CC the printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences

SQ Sequence 2572 BP; 762 A; 512 C; 566 G; 729 T; 0 U; 3 Other;

Query Match 97.4%; Score 2406.6; DB 6; Length 2572;
 Best Local Similarity 99.8%; Pred. NO. 0;
 Matches 2409; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 TATGAGCCTTCGNACTTGTGAGAGACTACAAAGTTTGGTGGTTATGTCCTTTAGT 60
 |||||
 Db 126 TATGAGCCTTCGNACTTGTGAGAGACTACAAAGTTTGGTGGTTATGTCCTTTAGT 185
 |||||
 Qy 61 TGGGCTCATACATTTGGGTTGGTACAGAAATCAAAAGCAGCCCTGTTTCCAAATACCTAA 120
 |||||
 Db 186 TGGGCTCATACATTTGGGTTGGTACAGAAATCAAAAGCAGCCCTGTTTCCAAATACCTAA 245
 |||||
 Qy 121 AAACGACGACATTCCTGACGACAGATAGTCTGGACCTTCAAAATCTTCAGAGAGCCAAAT 180
 |||||
 Db 246 AAACGACGACATTCCTGACGACAGATAGTCTGGACCTTCAAAATCTTCAGAGAGCCAAAT 305
 |||||
 Qy 181 CCAGGGGAAGTAGCAGGCTTGCAATCTTCAGGTAAAGAGCAGCTTTGAATCTGAGCTTC 240
 |||||
 Db 306 CCAGGGGAAGTAGCAGGCTTGCAATCTTCAGGTAAAGAGCAGCTTTGAATCTGAGCTTC 365
 |||||
 Qy 241 ATATCGAAAGAGAGATGAAATATACAGTTTGGATTTAGAAAGAACTGGCTTCTTTGAGCT 300
 |||||
 Db 366 ATATCGAAAGAGAGATGAAATATACAGTTTGGATTTAGAAAGAACTGGCTTCTTTGAGCT 425
 |||||
 Qy 301 GGGATATCTTTTATAGGTTCCATCTTGGAAATATCTTTTGCAGAGGTTGCAAAAGCAG 360
 |||||
 Db 426 GGGATATCTTTTATAGGTTCCATCTTGGAAATATCTTTTGCAGAGGTTGCAAAAGCAG 485
 |||||
 Qy 361 TCTGTAAATTTTCAGTCTCAGAGCAAAAGAGATTTTGAAGAGTGAAGTGAATTAAT 420
 |||||
 Db 486 TCTGTAAATTTTCAGTCTCAGAGCAAAAGAGATTTTGAAGAGTGAAGTGAATTAAT 545
 |||||
 Qy 421 ATTTGGAATTACTAATTTTGTCAATTAATCTATGCTGATTAGCTTCATAAAATTTGA 480
 |||||
 Db 546 ATTTGGAATTACTAATTTTGTCAATTAATCTATGCTGATTAGCTTCATAAAATTTGA 605
 |||||
 Qy 481 ACTTTTGAATTTTATAGCCACAATGCTGCATATTCATATCTTTAAATCTCTTAAAGATTAAT 540
 |||||
 Db 606 ACTTTTGAATTTTATAGCCACAATGCTGCATATTCATATCTTTAAATCTCTTAAAGATTAAT 665
 |||||
 Qy 541 TTTTAATGTTAAACCGTGATTAATGCAATAAATAGAAAATGTGGTTTACAAAATTAACACG 600
 |||||
 Db 666 TTTTAATGTTAAACCGTGATTAATGCAATAAATAGAAAATGTGGTTTACAAAATTAACACG 725
 |||||
 Qy 601 GTCTTCCTAGTTTACCACTGAAGTATGCTCGTTTGGAAAGCTTAAGAGCCATCAT 660
 |||||
 Db 726 GTCTTCCTAGTTTACCACTGAAGTATGCTCGTTTGGAAAGCTTAAGAGCCATCAT 785
 |||||
 Qy 661 GTGTAGAGTGAACCACTGACAACTGAGAGAGTCAAGGACCACACTTTTGTCTTGAAGAG 720
 |||||
 Db 786 GTGTAGAGTGAACCACTGACAACTGAGAGAGTCAAGGACCACACTTTTGTCTTGAAGAG 845
 |||||
 Qy 721 AATTGTACATCATCTATGCGCCCTCAGGTAGGCTGAAGCAGCTGCACAATGGCTTTGG 780
 |||||
 Db 846 AATTGTACATCATCTATGCGCCCTCAGGTAGGCTGAAGCAGCTGCACAATGGCTTTGG 905
 |||||
 Qy 781 AGGTTACGTGTGTACAACCTCACAGTCTCAGCTCTGCTCAGTCACTCTTTTGGTTCACACA 840
 |||||
 Db 906 AGGTTACGTGTGTACAACCTCACAGTCTCAGCTCTGCTCAGTCACTCTTTTGGTTCACACA 965
 |||||
 Qy 841 TCCCAATTTAAAGATCTTGACAGCCTCCATACAGAAATCATGTGTCAAGCTTCAGTGATTG 900
 |||||
 Db 966 TCCCAATTTAAAGATCTTGACAGCCTCCATACAGAAATCATGTGTCAAGCTTCAGTGATTG 1025
 |||||

QY	901	TGCTTATTACAGCTATTCTTTGCTGCAACCTGATTGAAAATGTTTCAGAGATTAGCCTT	960
DB	1026	TGCTTATTACAGCTATTCTTTGCTGCAACCTGATTGAAAATGTTTCAGAGATTAGCCTT	1085
QY	961	GACACCCACACCTGTCAATTAGATTAAATAAACAATCTTTTGAGTCTTTTGCAATCAGTTATCT	1020
DB	1086	GACACCCACACCTGTCAATTAGATTAAATAAACAATCTTTTGAGTCTTTTGCAATCAGTTATCT	1145
QY	1021	CAAGTCTGATACCTGTGGTGTGCAATCCAGTGGACTTTTAGTAGTACTCAGATCCCTCT	1080
DB	1146	CAAGTCTGAGACCTGTGGTGTGCAATCCAGTGGACTTTTAGTAGTACTCAGATCCCTCT	1205
QY	1081	TTCTTTGGTGGCTAGTATATAACAAGTAAACCTGCTGTATGCTCACCAAGAAAGAAAC	1140
DB	1206	TTCTTTGGTGGCTAGTATATAACAAGTAAACCTGCTGTATGCTCACCAAGAAAGAAAC	1265
QY	1141	AGAGCATGTCAAGTCTTTGATCCTCGAGAGCTTTTGTCTTACAAATCCAGAAAATCTGA	1200
DB	1266	AGAGCATGTCAAGTCTTTGATCCTCGAGAGCTTTTGTCTTACAAATCCAGAAAATCTGA	1325
QY	1201	AGGCCACATCAATTTAGGAAAGATTAAATTGATACCTTTAAAGGTCAAAGAGTTATAGA	1260
DB	1326	AGGCCACATCAATTTAGGAAAGATTAAATTGATACCTTTAAAGGTCAAAGAGTTATAGA	1385
QY	1261	TTCCACTGTATTACCTGGGATACCTCAATTGAAATGTGAGAGTTCAATTAATGAGGCTATT	1320
DB	1386	TTCCACTGTATTACCTGGGATACCTCAATTGAAATGTGAGAGTTCAATTAATGAGGCTATT	1445
QY	1321	ACCTATCAAAAAATCAACTGCCCTCAAGGTGGCACTCTTTGTACAACTTTATCCGGAGA	1380
DB	1446	ACCTATCAAAAAATCAACTGCCCTCAAGGTGGCACTCTTTGTACAACTTTATCCGGAGA	1505
QY	1381	CACCTTCGACACTGGAGAAGAACTGTGGTGGTCAGTTATGGGGTTTCTCTTGAAAATGC	1440
DB	1506	CACCTTCGACACTGGAGAAGAACTGTGGTGGTCAGTTATGGGGTTTCTCTTGAAAATGC	1565
QY	1441	AGTCTTGGACCACTCTTAACCTAGGAAGGCACTTAATCAGTGACCACTAGATCTTGT	1500
DB	1566	AGTCTTGGACCACTCTTAACCTAGGAAGGCACTTAATCAGTGACCACTAGATCTTGT	1625
QY	1501	CCTGTGCCAAAAAGTTATACATCCATCTTTTGAAGCAGTTTCTCAATATGATCATCGTATTAT	1560
DB	1626	CCTGTGCCAAAAAGTTATACATCCATCTTTTGAAGCAGTTTCTCAATATGATCATCGTATTAT	1685
QY	1561	TGCCATAGACAGAAATTGGAGTGACTCTGATGGAAACCCCTGACTAAAATGACAGGAACACA	1620
DB	1686	TGCCATAGACAGAAATTGGAGTGACTCTGATGGAAACCCCTGACTAAAATGACAGGAACACA	1745
QY	1621	GCCTATTGGATCCCTAGGCTCAATATGCTCTAATAGTTATGGAGTGTGAAAGATGTGTG	1680
DB	1746	GCCTATTGGATCCCTAGGCTCAATATGCTCTAATAGTTATGGAGTGTGAAAGATGTGTG	1805
QY	1681	CACGTCAAAAATTTGGCTCCCAACATTTTTTTCATCTTATTCTTAATGAAGCAACAATCTG	1740
DB	1806	CACGTCAAAAATTTGGCTCCCAACATTTTTTTCATCTTATTCTTAATGAAGCAACAATCTG	1865
QY	1741	CAGCTTGCTTCTCTGCAACAGAAATGACACTGCTGGGATGAGCTGAAGCTCACGTGTCA	1800
DB	1866	CAGCTTGCTTCTCTGCAACAGAAATGACACTGCTGGGATGAGCTGAAGCTCACGTGTCA	1925
QY	1801	GAGGCACTGCATGCTCTCGAGTTAACACTCAAGGAACCAATGGGCTTTGTTGGAGGTGG	1860
DB	1926	GAGGCACTGCATGCTCTCGAGTTAACACTCAAGGAACCAATGGGCTTTGTTGGAGGTGG	1985
QY	1861	CTGTACTGAAACTCATTTGGCTGCATATATCAGACACAGACTCACAAACGACCCAGAAAG	1920
DB	1986	CTGTACTGAAACTCATTTGGCTGCATATATCAGACACAGACTCACAAACGACCCAGAAAG	2045
QY	1921	CATTCTCAAGATGATGAATGATCTCAAAACAGAACTTCAATTAATTTGCTGAAGCATTTTG	1980
DB	2046	CATTCTCAAGATGATGAATGATCTCAAAACAGAACTTCAATTAATTTGCTGAAGCATTTTG	2105
QY	1981	CAGTGCCTAGAAATCTGTTTGGCTCTTTTAGAACATGATGAGGTGAAATTTCTCACTGA	2040
DB	2106	CAGTGCCTAGAAATCTGTTTGGCTCTTTTAGAACATGATGAGGTGAAATTTCTCACTGA	2165
QY	2041	CATGAAGTATGGACACCTTTTGGTCAAGTTCAGGCAGATTCTCCCTGTGTGCTAACTGGCC	2100
DB	2166	CATGAAGTATGGACACCTTTTGGTCAAGTTCAGGCAGATTCTCCCTGTGTGCTAACTGGCC	2225
QY	2101	AGATTTGCTTTCAAGTGTGGCTGTGGATTATACAAATAGCCAGGAAGAACTCAACTGGTC	2160
DB	2226	AGATTTGCTTTCAAGTGTGGCTGTGGATTATACAAATAGCCAGGAAGAACTCAACTGGTC	2285
QY	2161	TTTCTTAAAGAAGCACAGCTGCTCAATTTGTCGCCACAAAGCTGCTTCCACATGAAGCTGT	2220
DB	2286	TTTCTTAAAGAAGCACAGCTGCTCAATTTGTCGCCACAAAGCTGCTTCCACATGAAGCTGT	2345
QY	2221	GGGCTCAGCCAGCAACCTTGACCTTGGACTTTGTGCTGCAAAAGCTTAGTGGCTACAGGT	2280
DB	2346	GGGCTCAGCCAGCAACCTTGACCTTGGACTTTGTGCTGCAAAAGCTTAGTGGCTACAGGT	2405
QY	2281	GGCTGTAGAGACAGCCAAATTTGATTTGGGATCTTTTCATATGTTATTTGAAGATAAAACTA	2340
DB	2406	GGCTGTAGAGACAGCCAAATTTGATTTGGGATCTTTTCATATGTTATTTGAAGATAAAACTA	2465
QY	2341	AGAGAATAGCATGTTGCTATTACAGAGAAACAAATAAACTAGTCTGTTGGCAATTGAGA	2400
DB	2466	AGAGAATAGCATGTTGCTATTACAGAGAAACAAATAAACTAGTCTGTTGGCAATTGAGA	2525
QY	2401	AAAAAAAAAAAAAA 2413	
DB	2526	AAAAAAAAAAAAAA 2538	
RESULT 4			
AAA08567			
ID	AAA08567 standard; DNA; 2776 BP.		
XX	AC AAA08567;		
XX	DT 19-JUL-2000 (first entry)		
XX	Human chaperone protein 5 (HCHP-5) coding sequence.		
XX	Human chaperone protein; HCHP-5; neurodegenerative; metabolic;		
XX	developmental; autoimmune; inflammatory; cell proliferation; cancer; ss.		
XX	Homo sapiens.		
XX	Key Location/Qualifiers		
XX	CDS 748..2460		
XX	FT /*tag= a		
XX	FT /product= "HCHP-5"		
XX	WO200017358-A2.		
XX	30-MAR-2000.		
XX	22-SEP-1999; 99WO-US022027.		
XX	22-SEP-1998; 98US-00158642.		
XX	22-SEP-1998; 98US-0172221P.		
XX	19-JAN-1999; 99US-00233291.		
XX	19-APR-1999; 99US-00294698.		
XX	19-APR-1999; 99US-0172232P.		
XX	(INCY-) INCYTE PHARM INC.		
XX	Tang YT, Hillman JL, Yue H, Patterson C, Baughn MR, Batra S;		
XX	WPI; 2000-283583/24.		
XX	P-PSDB; AAY91944.		
XX	New purified polypeptides and polynucleotides encoding human chaperone		

PT	proteins, useful for diagnosing, treating and preventing disorders
PT	associated with the expression human chaperone proteins.
XX	
PS	Claim 7; Page 79-80; 88pp; English.
XX	
CC	AAA08563-68 encode human chaperone proteins 1-6 (HCHP-1 to HCHP-6)
CC	respectively. The sequences can be used to treat and prevent disorders
CC	associated with altered expression or activity of HCHP comprising
CC	administering a composition comprising the polypeptide or an antagonist
CC	to a patient (claimed). The human chaperone proteins are also useful for
CC	the diagnosis, treatment or prevention of neurodegenerative, metabolic,
CC	developmental, autoimmune/inflammatory disorders and cell proliferative
CC	disorders including cancer
XX	
SQ	Sequence 2776 BP; 834 A; 535 C; 593 G; 814 T; 0 U; 0 Other;
Query Match	97.1%; Score 2399.2; DB 3; Length 2776;
Best Local Similarity	99.9%; Pred. No. 0;
Matches 2401; Conservative	0; Mismatches 3; Indels 0; Gaps 0;
Qy	1 TATGAGCCTTCGGAACTTGTGGAGAGACTACAAAGTTTGGTTGTTATGGTCCCTTTAGT 60
Db	120 TATGAGCCTTCGGAACTTGTGGAGAGACTACAAAGTTTGGTTGTTATGGTCCCTTTAGT 179
Qy	61 TGGGCTCATACATTTGGGGTGGTACAGAAATCAAAAGCAGCCCTGTTTCCAAATACCTAA 120
Db	180 TGGGCTCATACATTTGGGGTGGTACAGAAATCAAAAGCAGCCCTGTTTCCAAATACCTAA 239
Qy	121 AAACGACGACATTCCTGAGCAAGATAGTCTGGGACCTTCAAAATCTTCAGAAAGCCAAAT 180
Db	240 AAACGACGACATTCCTGAGCAAGATAGTCTGGGACCTTCAAAATCTTCAGAAAGCCAAAT 299
Qy	181 CCAGGGGAAGTAGCAGGCTTGCATCTTCAGGTAAAGAACGACCTTTGAATCTGAGCTTC 240
Db	300 CCAGGGGAAGTAGCAGGCTTGCATCTTCAGGTAAAGAACGACCTTTGAATCTGAGCTTC 359
Qy	241 ATATCGAAAGAGAGATGAAATAATACAGTTGGATTAGAAAGAACTGGCTTCTTTGAGCT 300
Db	360 ATATCGAAAGAGAGATGAAATAATACAGTTGGATTAGAAAGAACTGGCTTCTTTGAGCT 419
Qy	301 GGGATATCTTCATAGTGTCCATCTTGGAAACATACCTTTTGCAGAGGCTCGAAAGCAG 360
Db	420 GGGATATCTTCATAGTGTCCATCTTGGAAACATACCTTTTGCAGAGGCTCGAAAGCAG 479
Qy	361 TCTGTAAATTTTCAGTCTCAAAGCAAAAGAGATTTGAAGAGTGAAGTAAATAAAT 420
Db	480 TCTGTAAATTTTCAGTCTCAAAGCAAAAGAGATTTGAAGAGTGAAGTAAATAAAT 539
Qy	421 ATTTGGAAATTAATAATTTGTCAATTAATCAATCTATGCTGATTAGCTTCATAAACATTGA 480
Db	540 ATTTGGAAATTAATAATTTGTCAATTAATCAATCTATGCTGATTAGCTTCATAAACATTGA 599
Qy	481 ACTTTTGTATTTATAGCCCAATGTGCAATTTATCATCTTTAATTTCTTAAGAAATAAT 540
Db	600 ACTTTTGTATTTATAGCCCAATGTGCAATTTATCATCTTTAATTTCTTAAGAAATAAT 659
Qy	541 TTTTAATGTTTAAACGTGATAATGCAATAAATAGAAAATGTGTTTACAAAATAAAAACG 600
Db	660 TTTTAATGTTTAAACGTGATAATGCAATAAATAGAAAATGTGTTTACAAAATAAAAACG 719
Qy	601 GTCTTCACTAGTTACCACTGAAGTAAAGATGTCTCGTTTGGAAAGCTAAGAACCATCAT 660
Db	720 GTCTTCACTAGTTACCACTGAAGTAAAGATGTCTCGTTTGGAAAGCTAAGAACCATCAT 779
Qy	661 GTGTAAAGATGAACCACTGACAACTGAGAGAGTCAGAACCACTTTCTGTCTTGAAG 720
Db	780 GTGTAAAGATGAACCACTGACAACTGAGAGAGTCAGAACCACTTTCTGTCTTGAAG 839
Qy	721 AATTGTAAACATCATGCTATGSCCCCTCAGTAGGCTGAAGCAGCTGCACAATGGCTTTG 780
Db	840 AATTGTAAACATCATGCTATGSCCCCTCAGTAGGCTGAAGCAGCTGCACAATGGCTTTG 899
Qy	781 AGGTTACGTGTGTACAACTTCACAGTCCCTCAGCTCTGCTCAGTCACTTTTGGTCACACA 840

Db	900 AGGTACGTGTGTACAACTTCACAGTCTCTGCTCAGTCACTTTTGGTCACACA 959
Qy	841 TCCCATTTTAAAGATCCCTGACAGCCTCCATACAGATCATGTGTCAAGCTTCAGTCAT 900
Db	960 TCCCATTTTAAAGATCCCTGACAGCCTCCATACAGAAATCATGTGTCAAGCTTCAGTCAT 1019
Qy	901 TGGCTTATTTACAGCTATTTCTTTGTGCAACCTGATTGAAAATGTTTCAGAGATTAGGCT 960
Db	1020 TGGCTTATTTACAGCTATTTCTTTGTGCAACCTGATTGAAAATGTTTCAGAGATTAGGCT 1079
Qy	961 GACACCCACATCTGTCATTTAGATTTAAATAAAACATCTTTTGGAGTCTTTGTCATGTTATCT 1020
Db	1080 GACACCCACATCTGTCATTTAGATTTAAATAAAACATCTTTTGGAGTCTTTGTCATGTTATCT 1139
Qy	1021 CAACTCTGATACCTGTGGTTGTCGAATCCAGTGGACCTTTAGTAGTACTCAGATCTCTCT 1080
Db	1140 CAACTCTGAGACCTGTGGTTGTCGAATCCAGTGGACCTTTAGTAGTACTCAGATCTCTCT 1199
Qy	1081 TTGTTTGGTGGCTAGTATATTAAACAAAGTAAACCTGCTGTATGCTCACACAGAAAGAAAC 1140
Db	1200 TTGTTTGGTGGCTAGTATATTAAACAAAGTAAACCTGCTGTATGCTCACACAGAAAGAAAC 1259
Qy	1141 AGAGCATGTCACTGCTTTGATCTCAGAGCCTTTTTCGTTTACAATTCAGAAAATGCTGA 1200
Db	1260 AGAGCATGTCACTGCTTTGATCTCAGAGCCTTTTTCGTTTACAATTCAGAAAATGCTGA 1319
Qy	1201 AGGCCACATCATTTTGGAAAGAGTTAATTTGCTTCTTAAAGGTCAAAGGTTATAGA 1260
Db	1320 AGGCCACATCATTTTGGAAAGAGTTAATTTGCTTCTTAAAGGTCAAAGGTTATAGA 1379
Qy	1261 TTCCACTGTATTAACCTGGGATCTCAATTGAAATGTCAAGAGTTCAATTAATGAGGCTATT 1320
Db	1380 TTCCACTGTATTAACCTGGGATCTCAATTGAAATGTCAAGAGTTCAATTAATGAGGCTATT 1439
Qy	1321 ACCTATCAAAAAATCAATGCGCTCAAGGTGGCACTCTTTTGTACAATTTTATCCGGAGA 1380
Db	1440 ACCTATCAAAAAATCAATGCGCTCAAGGTGGCACTCTTTTGTACAATTTTATCCGGAGA 1499
Qy	1381 CACTTCTGACACTGGAGAGGAACTGTGGTGTGCTAGTTATGGGTTCTCTTGGAAATGC 1440
Db	1500 CACTTCTGACACTGGAGAGGAACTGTGGTGTGCTAGTTATGGGTTCTCTTGGAAATGC 1559
Qy	1441 AGCTTGGACCACTGCTTTAACTTAGGAAGGAGCTTAATCAGTACAGCAGCTAGATCTTGT 1500
Db	1560 AGCTTGGACCACTGCTTTAACTTAGGAAGGAGCTTAATCAGTACAGCAGCTAGATCTTGT 1619
Qy	1501 CCGTGTGCCCCAAAAGTTTATACATCCATCTTTTGAAGCAGTTTCTCAATATCATCGTATTAT 1560
Db	1620 CCGTGTGCCCCAAAAGTTTATACATCCATCTTTTGAAGCAGTTTCTCAATATCATCGTATTAT 1679
Qy	1561 TGCCATAGACAGAAATGGAGTGACTCTGATGGAAACCCCTGACTTAAATGACAGGACACA 1620
Db	1680 TGCCATAGACAGAAATGGAGTGACTCTGATGGAAACCCCTGACTTAAATGACAGGACACA 1739
Qy	1621 GCCTATTGGATCCCTAGGCTCAATATGCTTAATAGTATGGAAGTGTGAAAGATGTGTG 1680
Db	1740 GCCTATTGGATCCCTAGGCTCAATATGCTTAATAGTATGGAAGTGTGAAAGATGTGTG 1799
Qy	1681 CACTGCAAAATTTGGCTCCCAACATTTTCTTCTTAATTAAGCAACAATCTG 1740
Db	1800 CACTGCAAAATTTGGCTCCCAACATTTTCTTCTTCTTAATTAAGCAACAATCTG 1859
Qy	1741 CAGCTTGTCTCTGCAACAGAAATGACACTGCTCCCTGGGATGAGCTCAAGCTCAAGTGTCA 1800
Db	1860 CAGCTTGTCTCTGCAACAGAAATGACACTGCTCCCTGGGATGAGCTCAAGCTCAAGTGTCA 1919
Qy	1801 GACGCACTGCACTGTCTGCACTTAACTCAAGGAACCATCGGCTTTGTTGGAGGCTG 1860
Db	1920 GACGCACTGCACTGTCTGCACTTAACTCAAGGAACCATCGGCTTTGTTGGAGGCTG 1979
Qy	1861 CTGTACTGAAACTCAATTTGGCTGCATATATACAGACCAAGACTCAACGACCCAGAAAG 1920

Db 707 AAATGTGCTTTACAAAATAAAACGGTCTTCACTAGTTACACCTGAAGTAAGATGTCTC 766
Qy 636 GTTTGGAAGCTTAAGAAGCCATCATTTGTGAAGAGTGAACCACTGACAACTGAGAGAGTCA 695
Db 767 GTTTGGAAGCTTAAGAAGCCATCATTTGTGAAGAGTGAACCACTGAGAGAGTCA 826
Qy 696 GGACCAACATTTCTGCTTGAAGAAGATTTGAATCATCATGCTATGAGCCCTCAGGTAGGC 755
Db 827 GGACCAACATTTCTGCTTGAAGAAGATTTGAATCATCATGCTATGAGCCCTCAGGTAGGC 886
Qy 756 TGAAGCAGCTGCACATGCTTTGGAGGTTACGTGTGTACAACCTCAGCTCCTCAGCTC 815
Db 887 TGAAGCAGCTGCACATGCTTTGGAGGTTACGTGTGTACAACCTCAGCTCCTCAGCTC 946
Qy 816 TGCTCAGTCACCTTTTGGTGCACATCCCATTTTAAAGATCCTGACAGCCCTCCATACAGA 875
Db 947 TGCTCAGTCACCTTTTGGTGCACATCCCATTTTAAAGATCCTGACAGCCCTCCATACAGA 1006
Qy 876 ATCATGTGTCAGCTTCAGTGATTTGGCTTATTTCAACAGCTATTTCTTGTGCTCAACCTGA 935
Db 1007 ATCATGTGTCAGCTTCAGTGATTTGGCTTATTTCAACAGCTATTTCTTGTGCTCAACCTGA 1066
Qy 936 TTGAAAATGTCAGAGATTAAGGCTTGACACCCACCACTGTCATTTAGATTAATAAATCAATC 995
Db 1067 TTGAAAATGTCAGAGATTAAGGCTTGACACCCACCACTGTCATTTAGATTAATAAATCAATC 1126
Qy 996 TTTTGTAGCTTTGCATCAGTATTTCTCAAGTCTCATACCTGCTGTGTCGAATCCCAAGTGG 1055
Db 1127 TTTTGTAGCTTTGCATCAGTATTTCTCAAGTCTCATACCTGCTGTGTCGAATCCCAAGTGG 1186
Qy 1056 ACTTTAGTAGTACTCAGATCCTCTTTTGTGCTGCTAGTATTAATAAAGATTAATAAATCAATC 1115
Db 1187 ACTTTAGTAGTACTCAGATCCTCTTTTGTGCTGCTAGTATTAATAAAGATTAATAAATCAATC 1246
Qy 1116 CCTGTATGCTCACAGAAAGGAAACAGAGATGTCAAGTCTGCTGCTGCTGCTGCTGCTGCTGCT 1175
Db 1247 CCTGTATGCTCACAGAAAGGAAACAGAGATGTCAAGTCTGCTGCTGCTGCTGCTGCTGCTGCT 1306
Qy 1176 TGCTTCAATTTCCAGAAAATGCTGAAGGCCACATCATTTTGAAGAAAGATTTAATTTGTAC 1235
Db 1307 TGCTTCAATTTCCAGAAAATGCTGAAGGCCACATCATTTTGAAGAAAGATTTAATTTGTAC 1366
Qy 1236 CTTTAAAGGCTCAAGAGTTATAGATTTCACTGTATTAACCTGGGATCTCAATTTGAATGT 1295
Db 1367 CTTTAAAGGCTCAAGAGTTATAGATTTCACTGTATTAACCTGGGATCTCAATTTGAATGT 1426
Qy 1296 CAGAAGTTCAATTAATGAGCTATTAACCTATCAAAAAATCAACTGCCCTCAAGGTGGCAC 1355
Db 1427 CAGAAGTTCAATTAATGAGCTATTAACCTATCAAAAAATCAACTGCCCTCAAGGTGGCAC 1486
Qy 1356 TCTTTTGTACAACCTTTATCCGAGACACTTC-TGACACTGGAGAGAACTGTGGTGTTC 1414
Db 1487 TCTTTTGTACAACCTTTATCCGAGACACTTCGTTGACACTGGAGAGAACTGTGGTGTTC 1546
Qy 1415 AGTTATGGGTTTCTCTTGAATAATGCAGTCTTGACACAGCTGTAACTAGGAGGCAG 1474
Db 1547 AGTTATGGGTTTCTCTTGAATAATGCAGTCTTGACACAGCTGTAACTAGGAGGCAG 1606
Qy 1475 CTAATCAGTGACCACTAGATCTTTGCTGTGCCCCAAAAAGTTATACATCATCTTTTGAAG 1534
Db 1607 CTAATCAGTGACCACTAGATCTTTGCTGTGCCCCAAAAAGTTATACATCATCTTTTGAAG 1666
Qy 1535 CAGTTTCTCAATATGAT-CGTATTAATGCGATAGACAGATTT-GGAGTACTCTGATGG 1592
Db 1667 CAGTTTCTCAATATGATCGGTATTAATGCGATAGACAGATTTGGAGTACTCTGATGG 1726
Qy 1593 AACCCC----TGACTAAAATGACAGGAACACAGCTTATTTGGATCCCTAGGCTCAATATGT 1648
Db 1727 GAACCCCTGGACTTAAAATGACAGGAACACAGCTTATTTGGATCCCTAGGCTCAATATGT 1786
Qy 1649 CCTAATAGTTATGGAAATGTGAAGATGTGTGCACTGCAAAAATTTGGCTCCCAACATTTT 1708

Db 1787 CCTAATAGTTATGGAAGTGTGAAAGATGTGTGCACTGCAAAAATTTGGCTCCAAACATTTT 1846
Qy 1709 TTTTCACTTTATTTCTTAATGAAGCACAATCTGAGCTTCTGCTCTCTGCAACAGAAATGAC 1768
Db 1847 TTTTCACTTTATTTCTTAATGAAGCACAATCTGAGCTTCTGCTCTCTGCAACAGAAATGAC 1906
Qy 1769 ACTGCTGGGATGAGCTGAAAGCTCAGCTGTGAGACGGCACTGCAATCTCTGCAAGTTAAACA 1828
Db 1907 ACTGCTGGGATGAGCTGAAAGCTCAGCTGTGAGACGGCACTGCAATCTCTGCAAGTTAAACA 1966
Qy 1829 CTCAGGAACCAATGGGCTTTTGGAGGTGGCTGTACTGAAACATCATTTGGCTGCATAT 1888
Db 1967 CTCAGGAACCAATGGGCTTTTGGAGGTGGCTGTACTGAAACATCATTTGGCTGCATAT 2026
Qy 1889 ATCAGACACAGACTCACAAACGACCCAGAAAGCATTTCTCAAGATGATGAATGATCTCAA 1948
Db 2027 ATCAGACACAGACTCACAAACGACCCAGAAAGCATTTCTCAAGATGATGAATGATCTCAA 2086
Qy 1949 ACAGAACTTTCAATTAATTTGCTGAAGCATTTTGCAGTGGCTGAGAAATCTGTTGCTGCTCT 2008
Db 2087 ACAGAACTTTCAATTAATTTGCTGAAGCATTTTGCAGTGGCTGAGAAATCTGTTGCTGCTCT 2146
Qy 2009 TTAGAACATGATGGAGGTGAAATTTCTC-ACTGACATGAAGTATGGACAACCTTTGGTTCAGT 2067
Db 2147 TTAGAACATGATGGAGGTGAAATTTCTCAACTGACATGAAGTATGGACAACCTTTGGTTCAGT 2206
Qy 2068 TCAGGCAGATTTCTCCC---TGTTGCTAACTGGCCAGATTTGCTTCAAGTGTGGCTG 2124
Db 2207 TCAGGCAGATTTCTCCCCTGGGTTGCTGAACCTGGCCAGATTTGCTTCAAGTGTGGCTG 2266
Qy 2125 TGGATTATACATAGCCAGAAAGAACTCAACTGGTCTTTCTTT-AAGAAACACACGCT-CGT 2182
Db 2267 TGGATTATACATAGCCAGAAAGAACTCAACTGGTCTTTCTTTAAGAAACACACGCTCCGT 2326
Qy 2183 CCAATTTGCTCCACAAAGCT--GCCTTCCATGAAGTGTGGGCTCAGCCA-GCAACCTG 2239
Db 2327 CCAATTTGCTCCACAAAGCTTGGCTTTTCCACATGAAGTGTGGGCTCAGCCAGCAACCTG 2386
Qy 2240 ACCTTTGGAC-TGTTTGACTGCAAAAGCTT---AGTGGCTTACAGTGGCTGTAGACACG 2295
Db 2387 ACCTTTGGACTTGTGGCTGCAAAAGCTTTAGTGGGCTTCAAGGTGGCTGTAGACACAGC 2446
Qy 2296 CAAATTTGAATTTGGGATCTTTT--CATATGTTATTGAAGATATAAACTAAGAGAAATAGCATG 2353
Db 2447 CAAATTTGAATTTGGGCTCTTTTTCATATTTGTTGAAGATATAAACTAAGAGAAATAGCATG 2506
Qy 2354 TTCTGTTATCAAGAGAAACAAATAAACTAGTCTGTGGCAATTTGAGAAAAAATAA 2409
Db 2507 TTCTGTTATCAAGAGAAACAAATAAACTAGTCTGTGGCAATTTGAGAAAAAATAA 2562

RESULT 6
AAC99959
ID AAC99959 standard; cdna; 1921 BP.
XX
AC AAC99959;
XX
DT 13-MAR-2001 (first entry)
XX
DE Human secreted protein gene 142 SEQ ID NO:152.
XX
KW Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;
KW antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;
KW cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;
KW fungicide; ophthalmological; gene therapy; pathological condition;
KW autoimmune disease; rheumatoid arthritis; hyperproliferative disorder;
KW neoplasm; cardiovascular disorder; cardiac arrest; cerebral ischaemia;
KW cerebrovascular disorder; angiogenesis; nervous system disorder;
KW Alzheimer's disease; infection; ocular disorder; corneal infection;
KW wound healing; skin aging; food additive; preservative; ss.
XX
OS Homo sapiens.
XX

PN WO200070042-A1.
XX 23-NOV-2000.
PD
XX 11-MAY-2000; 2000WO-US012788.
PF
XX 13-MAY-1999; 99US-0134068P.
PR
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX Rosen CA, Ruben SM, Moore PA, Young PE, Komatsoulis GA, Birse CE;
PI Duan RD, Florence KA, Soppet DR;
PI WPI: 2000-679828/66.
XX P-PSDB; AAB56218.
XX
XX Isolated nucleic acid molecule encoding a human secreted protein is used
PT in preventing, treating or ameliorating a medical condition.
PT
XX
PS Claim 1; Page 930-931; 1065pp; English.
XX
XX The polynucleotide sequences given in AAC99818 to AAC99977 encode the
CC human secreted proteins given in AAB56077 to AAB56362. Human secreted
CC proteins have activities based on the tissues and cells the genes are
CC expressed in. Examples of activities include: immunosuppressive;
CC antiarthritic; antirheumatic; antiproliferative; cytostatic; cardiant;
CC vasotropic; cerebroprotective; nootropic; neuroprotective; antibacterial;
CC virucide; fungicide; and ophthalmological. The human secreted
CC polynucleotides and proteins can be used to prevent, treat or ameliorate
CC a medical condition in e.g. humans, mice, rabbits, goats, horses, cats,
CC dogs, chickens or sheep. They are also used in diagnosing a pathological
CC condition or susceptibility to a pathological condition. Disorders which
CC are diagnosed or treated include autoimmune diseases e.g. rheumatoid
CC arthritis, hyperproliferative disorders e.g. neoplasms of the breast or
CC liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular
CC disorders e.g. cerebral ischaemia, angiogenesis, nervous system disorders
CC e.g. Alzheimer's disease, infections caused by bacteria, viruses and
CC fungi and ocular disorders e.g. corneal infection. The proteins can also
CC be used to aid wound healing and epithelial cell proliferation, to
CC prevent skin aging due to sunburn, to maintain organs before
CC transplantation, for supporting cell culture of primary tissues, to
CC regenerate tissues and in chemotaxis. The proteins can also be used as a
CC food additive or preservative to increase or decrease storage
CC capabilities. AAC99809 to AAC99817 and AAB56076 represent sequences used
CC in the exemplification of the present invention
XX
SQ Sequence 1921 BP; 565 A; 380 C; 414 G; 556 T; 0 U; 6 Other;

Query Match 71.3%; Score 1763.2; DB 3; Length 1921;
Best Local Similarity 99.6%; Pred. No. 1.6e-281;
Matches 1765; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 TATGAGCCTTCGGAACCTTGGGAGAGACTACAAAGTTTGGTTGGTTATGGTCCCTTAGT 60
DB 126 TATGAGCCTTCGGAACCTTGGGAGAGACTACAAAGTTTGGTTGGTTATGGTCCCTTAGT 185

QY 61 TGGGCTCATACATTTGGGGTGGTACAGATCAAAAGCAGCCCTGTTTCCAAATACCTAA 120
DB 186 TGGGCTCATACATTTGGGGTGGTACAGATCAAAAGCAGCCCTGTTTCCAAATACCTAA 245

QY 121 AAACGACGACATTCCTGAGCAAGATAGTCTGGGACTTTCAAATCTTCAGAAAGCCAAAT 180
DB 246 AAACGACGACATTCCTGAGCAAGATAGTCTGGGACTTTCAAATCTTCAGAAAGCCAAAT 305

QY 181 CCAGGGGAAGTAGCAGCTTGCATCTTCAGGTAAAGAGCAGCTTTGGAATCTGAGCTTC 240
DB 306 CCAGGGGAAGTAGCAGCTTGCATCTTCAGGTAAAGAGCAGCTTTGGAATCTGAGCTTC 365

QY 241 ATATCGAAAGACAGATGAATAATACAGTTGGAATAGAAAGACCTGGCTTCTTGTAGCT 300
DB 366 ATATCGAAAGACAGATGAATAATACAGTTGGAATAGAAAGACCTGGCTTCTTGTAGCT 425

QY 301 GGGATATCTTTCATAGGTGCCATCTTGGAAACATACCTTTTTGCGAGAGGTCTGCAAGCAG 360

DB 426 GGGATATCTTTCATAGGTGTCATCTTGGAAACATACCTTTTTCGAGAGGTCTGCAAGCAG 485
QY 361 TCTGTAATAATTTTCAGTCTCAAAAGCAAAAGAGATTTGAAAGAGTGAAGTAAAAATAAT 420
DB 486 TCTGTAATAATTTTCAGTCTCAAAAGCAAAAGAGATTTGAAAGAGTGAAGTAAAAATAAT 545
QY 421 ATTTGGAATTTACTAATTTTGTCAATTAATCATTTCTATGCTGATAGCTTCATAACATTGA 480
DB 546 ATTTGGAATTTACTAATTTTGTCAATTAATCATTTCTATGCTGATAGCTTCATAACATTGA 605
QY 481 ACTTTTTCGATTTTATAGCCACCAATGCTGCATATTCATACCTTTAAATTCCTTAAAGAAATAAT 540
DB 606 ACTTTTTCGATTTTATAGCCACCAATGCTGCATATTCATACCTTTAAATTCCTTAAAGAAATAAT 665
QY 541 TTTAATGTTTAAACCGTGATAATGCAATAAATAAGAAAAATGTCGTTTCAAAAATAAAAAAG 600
DB 666 TTTAATGTTTAAACCGTGATAATGCAATAAATAAGAAAAATGTCGTTTCAAAAATAAAAAAG 725
QY 601 GTCTTCACCTAGTTACCACTGAACTGAAGATGTCGTTTGGAGGCTTGAAGGCCATCATTT 660
DB 726 GTCTTCACCTAGTTACCACTGAACTGAAGATGTCGTTTGGAGGCTTGAAGGCCATCATTT 785
QY 661 GTTGAAGAGTGAACCACTGACCACTGAGAGAGTTCAGGACCAACATTTCTGTCTTGAAGAG 720
DB 786 GTTGAAGAGTGAACCACTGACCACTGAGAGAGTTCAGGACCAACATTTCTGTCTTGAAGAG 845
QY 721 AATTGTAACATCATGTCATGTCGCCCTCAGTGAAGTGAAGAGTGCACAAATGGCTTTGG 780
DB 846 AATTGTAACATCATGTCATGTCGCCCTCAGTGAAGTGAAGAGTGCACAAATGGCTTTGG 905
QY 781 AGGTTACGTTGTACAACTCAGCTCCTCAGCTCTGCTCAGCTCAGCTTTCGTCACACA 840
DB 906 AGGTTACGTTGTACAACTCAGCTCCTCAGCTCTGCTCAGCTCAGCTTTCGTCACACA 965
QY 841 TCCCATTTTAAAGATCTGACAGCCCTCCATACAGATCATGTCGTCAGCTTTCAGTGATTTG 900
DB 966 TCCCATTTTAAAGATCTGACAGCCCTCCATACAGATCATGTCGTCAGCTTTCAGTGATTTG 1025
QY 901 TGGCTTATTCACAGCTATTTCTTTGTCGCAACTGATTTGAAAATGTTTCAGAGATTAGGCTT 960
DB 1026 TGGCTTATTCACAGCTATTTCTTTGTCGCAACTGATTTGAAAATGTTTCAGAGATTAGGCTT 1085
QY 961 GACACCCACCACTGTCATTTAGATTAATAAACAATCTTTTGGCTCTTTCAGCTATCT 1020
DB 1086 GACACCCACCACTGTCATTTAGATTAATAAACAATCTTTTGGCTCTTTCAGCTATCT 1145
QY 1021 CAAGTCTGATACCTGTCGTTGTCGAATCCAGTGGACTTTTAGTAGTACTCAGATCCTCCT 1080
DB 1146 CAAGTCTGAGACTGTCGTTGTCGAATCCAGTGGACTTTTAGTAGTACTCAGATCCTCCT 1205
QY 1081 TTGTTTGGTGGCTAGTATTAATAAAGTAAACCTGCTGTATGCTCACAGAAAGAAAC 1140
DB 1206 TTGTTTGGTGGCTAGTATTAATAAAGTAAACCTGCTGTATGCTCACAGAAAGAAAC 1265
QY 1141 AGAGCATGTCAGTCTTTGATCCTCAGAGCCCTTTTGGCTTACAATTCAGAAATGCTCA 1200
DB 1266 AGAGCATGTCAGTCTTTGATCCTCAGAGCCCTTTTGGCTTACAATTCAGAAATGCTCA 1325
QY 1201 AGGCCACATCATTTTAGGAAAGAGTTTAATTTGTAACCTTTTAAAGGTCAGAGGTTATAGA 1260
DB 1326 AGGCCACATCATTTTAGGAAAGAGTTTAATTTGTAACCTTTTAAAGGTCAGAGGTTATAGA 1385
QY 1261 TTCCCATGTTATTCCTGGGATACCTCATTTGAAATGTCAGAGTTCAATTAATGAGGCTATT 1320
DB 1386 TTCCCATGTTATTCCTGGGATACCTCATTTGAAATGTCAGAGTTCAATTAATGAGGCTATT 1445
QY 1321 ACCTATCAAAAATCAACTGCCCTCAAGGTGCACCTCTTTTGTACAACCTTTATCCGGAGA 1380
DB 1446 ACCTATCAAAAATCAACTGCCCTCAAGGTGCACCTCTTTTGTACAACCTTTATCCGGAGA 1505
QY 1381 CACTTCTGACACTGGAGAGGAACCTGTCGTCGTCAGTTATGCGGTTTCTCTTTGAAATGTC 1440

Db 602 ACCACGTAGATCTTGCTGCTGCAAAAAGTTATACATCACTTTGAAGCAGTTTCTCA 661
Qy 1545 ATATGATCGTATATTTGCCATAGACAGAAATGGAGTACTCTGATGGAAACCCCTGACTA 1604
Db 662 ATATGATCGTATATTTGCCATAGACAGAAATGGAGTACTCTGATGGAAACCCCTGACTA 721
Qy 1605 AAATGACAGAAACACAGCCTATTTGGATCCCTAGGCTCAATATGCTCTTAATAGTTATGGAA 1664
Db 722 AAATGACAGAAACACAGCCTATTTGGATCCCTAGGCTCAATATGCTCTTAATAGTTATGGAA 781
Qy 1665 GTGTGAAGATGTGTGACCTGCAAAATTTGGCTCCCAACATTTTTTTCATCTTTATTCCTA 1724
Db 782 GTGTGAAGATGTGTGACCTGCAAAATTTGGCTCCCAACATTTTTTTCATCTTTATTCCTA 841
Qy 1725 ATGAAGCAACAATCTGCAGCTGCTCTCTGCAACAGAAATGACACTGCGTGGGATGAGC 1784
Db 842 ATGAAGCAACAATCTGCAGCTGCTCTCTGCAACAGAAATGACACTGCGTGGGATGAGC 901
Qy 1785 TGAAGCTCAGTGTGACAGGCACTGCATGCTCTGCAAGTAACTCAAGGAACCATGGG 1844
Db 902 TGAAGCTCAGTGTGACAGGCACTGCATGCTCTGCAAGTAACTCAAGGAACCATGGG 961
Qy 1845 CTTTGTGGGAGTGGCTGTACTGAAACTCATTTGGCTGCATATATCAGACACAGACTC 1904
Db 962 CTTTGTGGGAGTGGCTGTACTGAAACTCATTTGGCTGCATATATCAGACACAGACTC 1021
Qy 1905 ACAACGACCAGAAAGCATTTCTCAAGATGATGAATGTACTCAAAACAGAACTTCAATTA 1964
Db 1022 ACAACGACCAGAAAGCATTTCTCAAGATGATGAATGTACTCAAAACAGAACTTCAATTA 1081
Qy 1965 TTGCTGAAGCAATTTGCAAGTCCCTAGATCTGTTGTTGGCTCTTTAGAACATGATGGAG 2024
Db 1082 TTGCTGAAGCAATTTGCAAGTCCCTAGATCTGTTGTTGGCTCTTTAGAACATGATGGAG 1141
Qy 2025 GTGAATTTCTCACTGACATGAAGTATGGACACCTTTGGTCAGTTTCAGGCAGATTTCCCT 2084
Db 1142 GTGAATTTCTCACTGACATGAAGTATGGACACCTTTGGTCAGTTTCAGGCAGATTTCCCT 1201
Qy 2085 GTGTTGCTAACTGGCCAGATTTGCTTTTCAAGTGTGGCTGTGGAATATACAATPAGCCAG 2144
Db 1202 GTGTTGCTAACTGGCCAGATTTGCTTTTCAAGTGTGGCTGTGGAATATACAATPAGCCAG 1261
Qy 2145 AAGAACTCACTGGTCTTTTAAAGAGCACAGTGTGTCATTTTGGCCACAAGCTGCC 2204
Db 1262 AAGAACTCACTGGTCTTTTAAAGAGCACAGTGTGTCATTTTGGCCACAAGCTGCC 1321
Qy 2205 TTCCACATGAAGCTGTGGGCTCAGCCAGCAACCTGACCTTGGACTGTTGACTGCAAAAGC 2264
Db 1322 TTCCACATGAAGCTGTGGGCTCAGCCAGCAACCTGACCTTGGACTGTTGACTGCAAAAGC 1381
Qy 2265 TTAGTGGCTTACAGTGGCTGTAGAGACAGCCAAATTTGATTTGGGATCTTTTCATATGTTA 2324
Db 1382 TTAGTGGCTTACAGTGGCTGTAGAGACAGCCAAATTTGATTTGGGATCTTTTCATATGTTA 1441
Qy 2325 TTGAAGATAAACTAAGAAATAGCATGTTCGTATTAACAGAAACAATAAATAGT 2384
Db 1442 TTGAAGATAAACTAAGAAATAGCATGTTCGTATTAACAGAAACAATAAATAGT 1501
Qy 2385 CTGTTGGCAATTCAGAAAAAATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 2419
Db 1502 CTGTTGGCAATTCAGAAAAAATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 1536

RESULT 8
ADP43721

ID ADP43721 standard; cDNA; 1546 BP.

AC ADP43721;

XX 18-NOV-2004 (first entry)

DT Human PMMM-37 encoding cDNA SEQ ID NO:78.

XX KW human; protein modification and maintenance molecule; PMMM;
KW gastrointestinal; cardiovascular; immunosuppressive; anti-inflammatory;
KW cytostatic; neuroprotective; gynaecological; gene therapy;
KW gastrointestinal disorder; cardiovascular disorder; autoimmune disorder;
KW inflammatory disorder; cell proliferative disorder;
KW developmental disorder; epithelial disorder; neurological disorder;
KW reproductive disorder; gene; ss; single nucleotide polymorphism; SNP.
XX Homo sapiens.
XX Key Location/Qualifiers
FT CDS 14..1510
FT /tag= a
FT /product= "PMMM-37"
FT variation replace(541,C)
FT /tag= b
FT variation /standard_name= "single nucleotide polymorphism (SNP)"
FT replace(547,T)
FT /tag= c
FT /standard_name= "single nucleotide polymorphism (SNP)"
XX WO2004053068-A2.
XX 24-JUN-2004.
XX 03-DEC-2003; 2003WO-US038573.
XX 05-DEC-2002; 2002US-0431639P.
XX 17-DEC-2002; 2002US-0434315P.
XX 24-JAN-2003; 2003US-0442442P.
XX 30-JAN-2003; 2003US-0444141P.
XX 21-FEB-2003; 2003US-0449491P.
XX (INCY-) INCYTE CORP.
XX Becha SD, Hafalia AJA, Swarnakar A, Ramkumar J, Richardson TW;
PI Kable AB, Marquis JP, Khare R, Lee SY, Tran UK, Bhatia UG;
PI Burdill JD, Blake JU, Ho A, Zheng W, Gao J, Chawla NK, Mason PM;
PI Jin P, Lee S;
XX WPI; 2004-468843/44.
XX P-PSDB; ADP43680.
XX New human protein modification and maintenance molecules and
XX polynucleotides for diagnosing, preventing or treating diseases
XX associated with aberrant protein expression, e.g. cardiovascular or cell
XX proliferative disorders.
XX Claim 5; SEQ ID NO 78; 268pp; English.
XX The present sequence encodes a human protein modification and maintenance
XX molecule (PMMM) (1), designated PMMM-37. Also described: (1) an isolated
XX polynucleotide encoding (1); (2) a recombinant polynucleotide comprising
XX a promoter sequence operably linked to the polynucleotide in (1); (3) a
XX cell transformed with the recombinant polynucleotide; (4) a transgenic
XX organism comprising the recombinant polynucleotide; (5) methods of
XX producing or purifying (1); (6) an isolated antibody that specifically
XX binds to (1); (7) detecting a target polynucleotide or (1) in a sample;
XX (8) compositions comprising the polypeptide, an agonist compound, an
XX antagonist compound or an antibody, and an excipient; (9) treating
XX diseases or conditions associated with decreased expression or
XX overexpression of functional human PMMM; (10) screening for a compound
XX that is effective as an agonist or antagonist of (1), that specifically
XX binds to (1), that modulates the activity of (1), or is effective in
XX altering expression of the target polynucleotide; (11) screening for
XX potential toxicity of a test compound; (12) a diagnostic test for a
XX condition or disease associated with the expression of PMMM in a
XX biological sample; (13) diagnosing a condition or disease associated with
XX the expression of PMMM in a subject; (14) preparing a polyclonal or
XX monoclonal antibody with the specificity of the antibody in (6); (15) a
XX polyclonal or monoclonal antibody produced by the method in (14); (16)
XX compositions comprising the polyclonal or monoclonal antibody, and a

CC carrier; (17) generating an expression profile of a sample containing the
 CC polynucleotides; and (18) an array comprising different nucleotide
 CC molecules affixed at distinct physical locations on a solid substrate,
 CC where at least one nucleotide molecule comprises a first oligonucleotide
 CC or polynucleotide sequence specifically hybridizable with at least 30
 CC contiguous nucleotides of the target polynucleotide. PWM sequences have
 CC gastrointestinal, cardiovascular, immunosuppressive, antiinflammatory,
 CC cytostatic, neuroprotective and gynaecological activities, and can be
 CC used in gene therapy. The composition and methods are useful for
 CC diagnosing, preventing or treating diseases or conditions associated with
 CC aberrant expression of PWM, such as gastrointestinal, cardiovascular,
 CC autoimmune/inflammatory, cell proliferative, developmental, epithelial,
 CC neurological or reproductive disorders. They may also be used for
 CC assessing the effects of exogenous compounds on the expression of nucleic
 CC acid and amino acid sequences of PWM. The PWM or its fragments are also
 CC useful in screening compounds for effectiveness as agonist or antagonist
 CC of the polypeptides, or in altering the expression of the target
 CC polynucleotide and compounds that specifically bind to or modulate the
 CC activity of the polypeptide. The microarray is useful in monitoring or
 CC measuring protein-protein interactions, drug-target interactions, and
 CC gene expression profiles.

XX Sequence 1546 BP; 445 A; 327 C; 331 G; 443 T; 0 U; 0 Other;

Query Match 53.2%; Score 1314.2; DB 13; Length 1546;
 Best Local Similarity 87.6%; Pred. No. 1.6e-207;
 Matches 1542; Conservative 0; Mismatches 3; Indels 216; Gaps 1;

QY 617 ACCTGAAGTAAGATGTCGTGTTGGAGCTGAAGAGCCATCATTTGTTGAAGTGGAACCA 676
 DB 2 ACCTGAAGTAAGATGTCGTGTTGGAGCTGAAGAGCCATCATTTGTTGAAGTGGAACCA 61
 QY 677 CTGACAACTGAGAGAGTCAGGACCACTTCTGCTCTGAAAGAAATGTTAACATCATGC 736
 DB 62 CTGACAACTGAGAGAGTCAGGACCACTTCTGCTCTTGAAGAAATGTTAACATCATGC 121
 QY 737 TATGGCCCTCAGGTAGGCTGAAGCAGCTGCACATGGCTTTGGAGGTTACGTTGTACA 796
 DB 122 TATGGCCCTCAGGTAGGCTGAAGCAGCTGCACATGGCTTTGGAGGTTACGTTGTACA 181
 QY 797 ACCTCAAGTCCTCAGCTGTCAGTCACTCTTTGGTGCACATCCCATTTTAAAGATC 856
 DB 182 ACCTCAAGTCCTCAGCTGTCAGTCACTCTTTGGTGCACATCCCATTTTAAAGATC 241
 QY 857 CTGACAGCTCCATACAGATCATGTCTCAAGCTTCAGTGTGTCATTCACAGCT 916
 DB 242 CTGACAGCTCCATACAGATCATGTCTCAAGCTTCAGTGTGTCATTCACAGCT 301
 QY 917 ATTCTTTGCTGCAACCTGATTTGAAAATGTTTCAGAGATTAGGCTTGACACCCACCTGTC 976
 DB 302 ATTCTTTGCTGCAACCTGATTTGAAAATGTTTCAGAGATTAGGCTTGACACCCACCTGTC 361
 QY 977 ATTAGATTAAATAAATCTTTTGAAGTCTTTTGATCAGTTATCTCAAGTCTGATACCTGT 1036
 DB 362 ATTAGATTAAATAAATCTTTTGAAGTCTTTTGATCAGTTATCTCAAGTCTGATACCTGT 421
 QY 1037 GGTGTGCGAATCCAGTGACCTTTAGTACTCTCAGATCCTCTTTGTTGGTGGTAGT 1096
 DB 422 GGTGTGCGAATCCAGTGACCTTTAGTACTCTCAGATCCTCTTTGTTGGTGGTAGT 481
 QY 1097 ATATTAAACAAGTAAACCTGCTGTATGCTCACAGAAAGAAACAGAGCATGTCAAGTGCT 1156
 DB 482 ATATTAAACAAGTAAACCTGCTGTATGCTCACAGAAAGAAACAGAGCATGTCAAGTGCT 541
 QY 1157 TTGATCCTGAGAGCCCTTTTGGCTTACAAATTCAGAAAAATGCTGAAGGCCACATCAATTTTA 1216
 DB 542 TTGATCCTGAGAGCCCTTTTGGCTTACAAATTCAGAAAAATGCTGAAGGCCACATCAATTTTA 601
 QY 1217 GGAAGAGTTAATTTGTAAGTTTAAAGGTCAGAGATTATAGATTCACCTGTATTAACCT 1276
 DB 602 GGAAGAGTTAATTTGTAAGTTTAAAGGTCAGAGATTATAGATTCACCTGTATTAACCT 661
 QY 1277 GGGATACCTCATTTGAAAATGTGAGAAGTTCAATTAATGAGGCTATTACCTATCAAAAAATCA 1336

DB 662 GGGATACCTCATTTGAAAATGTGAGAAGTTCAATTAATGAGGCTATTACCTATCAAAAAATCA 721
 QY 1337 ACTGCCCTCAAGGTGGCACTCTTTTGTACAACTTTATCCGGAGACACTTCTGACACTGGA 1396
 DB 722 ACTGCCCTCAAGGTGGCACTCTTTTGTACAACTTTATCCGGAGACACTTCTGACACTGGA 781
 QY 1397 GAAGGAACCTGTCGTGTCAGTTATGGGGTTCTCTTGAATAATGCACTCTTGGACCAAGCTG 1456
 DB 782 GAAGGAACCTGTCGTGTCAGTTATGGGGTTCTCTTGAATAATGCACTCTTGGACCAAGCTG 841
 QY 1457 CTTAAACCTAGGAAGCAGCTAATCAGTGAACCACTGATGTCCTCTGTGCCAAAAGTT 1516
 DB 842 CTTAAACCTAGGAAGCAGCTAATCAGTGAACCACTGATGTCCTCTGTGCCAAAAGTT 901
 QY 1517 ATACATCCATCTTTGAAGCAGTTTCTCAATATGATCGTATTTATTTGCCATAGACAAAT 1576
 DB 902 ATACATCCATCTTTGAAGCAGTTTCTCAATATGATCGTATTTATTTGCCATAGACAAAT 961
 QY 1577 GGAGTGACTCTGATGGAACCCCTGACTAAAATGACAGGAACACAGCTATTGGATCCCTA 1636
 DB 962 GGAGTGACTCTGATGGAACCCCTGACTAAAATGACAGGAACACAGCTATTGGATCCCTA 1021
 QY 1637 GGTCAATATGTCCTAATAGTTATGGAAGTGTGAAAGATGTGTGCACTGCAAAAATTTGTC 1696
 DB 1022 GGTCAATATGTCCTAATAGTTATGGAAGTGTGAAAGATGTGTGCACTGCAAAAATTTGTC 1081
 QY 1697 TCCCAACATTTTTTTCATCTTATCTTAATGAAGCAACATCTGACAGCTTGTCTCTGC 1756
 DB 1082 TCCCAACATTTTTTTCATCTTATCTTAATGAAGCAACATCTGACAGCTTGTCTCTGC 1141
 QY 1757 AACAGAAATGACACTGCGCTGGGATGAGCTGAAGCTCACGTGTCAGACGGCACTGCAATGC 1816
 DB 1142 AACAGAAATGACACTGCGCTGGGATGAGCTGAAGCTCACGTGTCAGACGGCACTGCAATGC 1201
 QY 1817 CTCGAGTTAAACCTCAAGGAACCATGGGCTTTGTTGGAGGTGGCTGTACTGAAACTCAT 1876
 DB 1202 CTCGAGTTAAACCTCAAGGAACCATGGGCTTTGTTGGAGGTGGCTGTACTGAAACTCAT 1261
 QY 1877 TTGGCTGCATATATACAGACACAAAGACTCACAAAGCCCAAGAAAGCATTTCTCAAAGATGAT 1936
 DB 1262 TTGGCTGCATATATACAGACACAAAGACTCACAAAGCCCAAGAAAGCATTTCTCAAAGATGAT 1283
 QY 1937 GAATGCTACTCAACAGAACTTCAATTAATGCTGAAGCATTTTTCAGAGTGCCTTAGAATCT 1996
 DB 1284 ----- 1283
 QY 1997 GTTGTGGCTCTTTAGAACATGATGGAGGTGAAATTTCTCACTGACATGAAGTATGGACAC 2056
 DB 1284 ----- 1283
 QY 2057 CTTTGGTCAGTTACGGCAGATTCTCCCTGTGTGTGCTTAAGTGGCCAGATTGCTTTTACAG 2116
 DB 1284 -----AG 1285
 QY 2117 TGTGCTGTGGATTATACAAATAGCCAGGAAGAACTCAACTGGTCTTTCTTAAGAAAGCACA 2176
 DB 1286 TGTGCTGTGGATTATACAAATAGCCAGGAAGAACTCAACTGGTCTTTCTTAAGAAAGCACA 1345
 QY 2177 CGTCTGCAATTTGTGCCCAAAAGCTGCGCTTCCACATGAAGCTGTGGGCTCAGCCAGCAAC 2236
 DB 1346 CGTCTGCAATTTGTGCCCAAAAGCTGCGCTTCCACATGAAGCTGTGGGCTCAGCCAGCAAC 1405
 QY 2237 CTGACCTTGGACTGTTTGACTGCAAGCTTAGTGCCCTACAGGTGGCTGTAGAGACAGCC 2296
 DB 1406 CTGACCTTGGACTGTTTGACTGCAAGCTTAGTGCCCTACAGGTGGCTGTAGAGACAGCC 1465
 QY 2297 AATTTGATTTGGGATCTTTTCATATCTTATGAAGTAAAACTAAGAGAAATAGCATGTTC 2356
 DB 1466 AATTTGATTTGGGATCTTTTCATATCTTATGAAGTAAAACTAAGAGAAATAGCATGTTC 1525
 QY 2357 GTATTACAAGAGAAAAAATA 2377

Db 1526 GTATTACAGAGAAACAATA 1546

RESULT 9
AAC99977

ID AAC99977 standard; cDNA; 1274 BP.

AC AAC99977;

XX 13-MAR-2001 (first entry)

DT Human secreted protein gene 142 SEQ ID NO:170.

XX Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;
XX antirheumatic; antiproliferative; cytostatic; cardiac; vasotropic;
XX cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;
XX fungicide; ophthalmologic; gene therapy; pathological condition;
XX autoimmune disease; rheumatoid arthritis; hyperproliferative disorder;
XX neoplasm; cardiovascular disorder; cardiac arrest; cerebral ischaemia;
XX cerebrovascular disorder; angiogenesis; nervous system disorder;
XX Alzheimer's disease; infection; ocular disorder; corneal infection;
XX wound healing; skin aging; food additive; preservative; chromosome 20;
XX ss.

OS Homo sapiens.

XX WO200070042-A1.

PN 23-NOV-2000.

XX 11-MAY-2000; 2000WO-US012788.

XX 13-MAY-1999; 99US-0134068P.

PR (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM, Moore PA, Young PE, Komatsoulis GA, Birse CE;
XX Duan RD, Florence KA, Soppet DR;

XX WPI; 2000-679828/66.

DR P-PSDB; AAB56236.

XX Isolated nucleic acid molecule encoding a human secreted protein is used
XX in preventing, treating or ameliorating a medical condition.

PS Claim 1; Page 939-940; 1065pp; English.

XX The polynucleotide sequences given in AAC99818 to AAC99977 encode the
CC human secreted proteins given in AAB56077 to AAB56362. Human secreted
CC proteins have activities based on the tissues and cells the genes are
CC expressed in. Examples of activities include: immunosuppressive;
CC antiarthritic; antirheumatic; antiproliferative; cytostatic; cardiac;
CC vasotropic; cerebroprotective; nootropic; neuroprotective; antibacterial;
CC virucide; fungicide; and ophthalmologic. The human secreted
CC polynucleotides and proteins can be used to prevent, treat or ameliorate
CC a medical condition in e.g. humans, mice, rabbits, goats, horses, cats,
CC dogs, chickens or sheep. They are also used in diagnosing a pathological
CC condition or susceptibility to a pathological condition. Disorders which
CC are diagnosed or treated include autoimmune diseases e.g. rheumatoid
CC arthritis, hyperproliferative disorders e.g. neoplasms of the breast or
CC liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular
CC disorders e.g. cerebral ischaemia, angiogenesis, nervous system disorders
CC e.g. Alzheimer's disease, infections caused by bacteria, viruses and
CC fungi and ocular disorders e.g. corneal infection. The proteins can also
CC be used to aid wound healing and epithelial cell proliferation, to
CC prevent skin aging due to sunburn, to maintain organs before
CC transplantation, for supporting cell culture of primary tissues, to
CC regenerate tissues and in chemotaxis. The proteins can also be used as a
CC food additive or preservative to increase or decrease storage
CC capabilities. AAC99809 to AAC99817 and AAB56076 represent sequences used
CC in the exemplification of the present invention
XX Sequence 1274 BP; 390 A; 256 C; 271 G; 357 T; 0 U; 0 Other;

XX PF 17-DEC-1998; 98WO-IB0002122.
XX
XX 17-DEC-1997; 97US-0069957P.
PR 09-FEB-1998; 98US-0074121P.
PR 13-APR-1998; 98US-0081563P.
PR 10-AUG-1998; 98US-0096116P.
XX
XX (GEST) GENSET.
XX
XX Bougueleret L, Duclert A, Dumas Milne Edwards J;
PI
XX
XX WPI: 1999-385906/32.
DR P-PSDB; AAY35896.
XX
XX New isolated human secreted proteins.
PT
XX
XX Claim 1; Page 179; 516pp; English.
XX
XX This sequence represents an extended human secreted protein coding
CC sequence of the invention. The secreted proteins can be used in treating
CC or controlling a variety of human conditions. The secreted proteins may
CC act as cytokines or may affect cellular proliferation or differentiation
CC or may act as immune system regulators, haematopoiesis regulators, tissue
CC growth regulators, regulators of reproductive hormones or cell movement
CC or have chemotactic/chemokinetic, receptor/ligand, anti-inflammatory or
CC tumour inhibition activity. The DNAs can be used in forensic procedures
CC to identify individuals or in diagnostic procedures to identify
CC individuals having genetic diseases resulting from abnormal expression of
CC the genes corresponding to the extended cDNAs. They are also useful for
CC constructing a high resolution map of the human chromosomes. They can
CC also be used for gene therapy to control or treat genetic diseases
XX
XX Sequence 458 BP; 160 A; 74 C; 102 G; 121 T; 0 U; 1 Other;

Query Match 17.1%; Score 422.4; DB 2; Length 458;
Best Local Similarity 99.5%; Pred. No. 1.5e-60;
Matches 434; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
QY 1 TATGAGCCTTCGGAACCTTGTGGAGAGACTACAAAGTTTGGTGTATGTCCTTTAGT 60
DB 11 TATGAGCCTTCGGAACCTTGTGGAGAGACTACAAAGTTTGGTGTATGTCCTTTAGT 70
QY 61 TGGGCTCATACATTGGGGTGGTACAGATCAAAAGAGCCCTGTTTCCAAATACCTAA 120
DB 71 TGGGCTCATACATTGGGGTGGTACAGATCAAAAGAGCCCTGTTTCCAAATACCTAA 130
QY 121 AAACGACGACATTCCTGAGCAAGATAGTCGGGACTTTCAAATCTTCAGAAGGCCAAAT 180
DB 131 AAACGACGACATTCCTGAGCAAGATAGTCGGGACTTTCAAATCTTCAGAAGGCCAAAT 190
QY 181 CCAGGGGAAG-TAGCAGGCTTGCATCTTCAGGTAAGAGCAGCTTGAATCTGAGCTT 239
DB 191 CCAGGGGAAGTAGCAGGCTTGCATCTTCAGGTAAGAGCAGCTTGAATCTGAGCTT 250
QY 240 CATATCGAAGAGAGATGAAAAATACAGTTGGATTAGAAAGAACTGGCTTCTTTAGC 299
DB 251 CATATCGAAGAGAGATGAAAAATACAGTTGGATTAGAAAGAACTGGCTTCTTTAGC 310
QY 300 TGGGATATCTTTTCATAGGTGTCATCTTGGAAACATCTTTTGCAGAGGTCTGCAAGCA 359
DB 311 TGGGATATCTTTTCATAGGTGACCATCTTGGAAACATCTTTTGCAGAGGTCTGCAAGCA 370
QY 360 GTCTGTAATAATTTTCAGTCTCAAAAGCAAAACAAAGAGATTTGAAGAGTGAAGTAAATAAA 419
DB 371 GTCTGTAATAATTTTCAGTCTCAAAAGCAAAACAAAGAGATTTGAAGAGTGAAGTAAATAAA 430
QY 420 TATTGGAATTACTAA 435
DB 431 TATTGGAATTACTAA 446

ADP18847
ID ADP18847 standard; cDNA; 458 BP.
XX
XX ADP18847;
AC
XX 26-AUG-2004 (first entry)
DT
XX Human secreted polynucleotide #103.
DE
XX Human; secreted protein; gene; ss; genetic disease.
KW
XX Homo sapiens.
OS
XX US2004110939-A1.
PN
XX 10-JUN-2004.
PD
XX 15-OCT-2001; 2001US-00978360.
PF
XX 17-DEC-1998; 98WO-IB0002122.
PR 09-FEB-1999; 99WO-IB000282.
PR 21-JUN-2000; 2000WO-IB000951.
PR 15-SEP-2000; 2000US-00663600.
XX
XX (GEST) GENSET SA.
XX
XX Dumas Milne Edwards J, Bougueleret L, Jobert S, Clusel C;
PI Duclert A;
XX
XX WPI: 2004-440404/41.
DR P-PSDB; ADP19252.
XX
XX New isolated polynucleotide encoding secreted polypeptide, useful for
PT gene therapy, or in diagnostic procedures to identify individuals having
PT genetic diseases resulting from abnormal expression of the genes.
XX
XX Claim 1; SEQ ID NO 103; 113pp; English.
XX
XX The invention relates to human cDNA sequences that encode human secreted
CC proteins. The invention also relates to an antibody that specifically
CC binds to a polypeptide of the invention and a method of binding the
CC polypeptide to an antibody. The polynucleotides are useful for expressing
CC the entire secreted proteins which they encode and for distinguishing
CC human tissues and cells from non-human tissues and cells, and for
CC distinguishing between human tissues and cells that do or do not express
CC the polynucleotides comprising the cDNAs. The polynucleotides and
CC polypeptides are useful in forensic procedures or diagnostic procedures
CC to identify individuals with genetic diseases resulting from abnormal
CC expression of the genes corresponding to the cDNAs. The sequences are
CC also useful in gene therapy to control or treat genetic diseases. This
CC sequence represents a human secreted polynucleotide of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format from USPTO at
CC seqdata.uspto.gov/sequence.html.
XX
XX Sequence 458 BP; 160 A; 74 C; 102 G; 121 T; 0 U; 1 Other;

Query Match 17.1%; Score 422.4; DB 12; Length 458;
Best Local Similarity 99.5%; Pred. No. 1.5e-60;
Matches 434; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
QY 1 TATGAGCCTTCGGAACCTTGTGGAGAGACTACAAAGTTTGGTGTATGTCCTTTAGT 60
DB 11 TATGAGCCTTCGGAACCTTGTGGAGAGACTACAAAGTTTGGTGTATGTCCTTTAGT 70
QY 61 TGGGCTCATACATTGGGGTGGTACAGATCAAAAGAGCCCTGTTTCCAAATACCTAA 120
DB 71 TGGGCTCATACATTGGGGTGGTACAGATCAAAAGAGCCCTGTTTCCAAATACCTAA 130
QY 121 AAACGACGACATTCCTGAGCAAGATAGTCGGGACTTTCAAATCTTCAGAAGGCCAAAT 180
DB 131 AAACGACGACATTCCTGAGCAAGATAGTCGGGACTTTCAAATCTTCAGAAGGCCAAAT 190

PR 28-AUG-2000; 2000US-00649811.
XX (CORI-) CORIXA CORP.
XX
XX Xu J, Lodes MJ, Secretist H, Benson DR, Meagher MJ, Stolk JA;
PI King GE, Wang T, Jiang Y;
XX
XX WPI; 2001-441847/47.
XX
XX
XX Colon tumor associated proteins and nucleic acids useful for the
PT prevention, diagnosis and treatment of colonic cancer.
XX
XX Claim 25; Page 174; 472pp; English.
XX
XX The present invention describes colon tumour associated proteins (I) and
CC the polynucleotides (II) that encode them. (I) have cytostatic activity.
CC (I) and (II) can be used in gene therapy and vaccine production. (I) and
CC (II) may be used in the prevention, diagnosis and treatment of diseases
CC associated with inappropriate colon tumour associated protein (TCAP)
CC expression, such as colonic cancer. For example, (I) and (II) may be used
CC to treat disorders associated with decreased expression by rectifying
CC mutations or deletions in a patient's genome that affect the activity of
CC TCAPs by expressing inactive proteins or to supplement the patients own
CC production of them. Additionally, (II) may be used to produce the TCAP
CC proteins, by inserting the nucleic acids into a host cell culturing the
CC cell to express the protein. (II) and its complementary sequences may
CC also be used as DNA probes in diagnostic polymerase chain reaction (PCR)
CC and hybridisation assays to detect and quantitate the presence of similar
CC nucleic acids in samples, and therefore which patients may be in need of
CC restorative therapy. (I) may also be used as antigens in the production
CC of antibodies against TCAPs and in assays to identify modulators of TCAP
CC expression and activity. Anti-(I) antibodies and antagonists may also be
CC used to down regulate TCAP expression and activity. The anti-(I)
CC antibodies may also be used as diagnostic agents for detecting the
CC presence of TCAPs in samples (e.g. by enzyme linked immunosorbant assay
CC (ELISA)). AA128460 to AA129512 and AA24494 to AA24523 represent
CC nucleotide and amino acid sequences given in the exemplification of the
CC present invention
XX
XX Sequence 421 BP; 113 A; 100 C; 91 G; 116 T; 0 U; 1 Other;

Query Match 16.7%; Score 413.4; DB 4; Length 421;
Best Local Similarity 99.5%; Pred. No. 4.7e-59;
Matches 414; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1745 TTGCTTCTCTGCAACAGAAATGACACTGCTGGATGAGCTGAAGCTCAGTGTCAAGC 1804
DB 416 TGGCTTCTCTGCAACAGAAATGACACTGCTGGATGAGCTGAAGCTCAGTGTCAAGC 357
QY 1805 GCACCTGCATGCTCTGCAAGTAACTAACACTCAAGGAACCAATGGCTTTGTTGGAGGTGGCTGT 1864
DB 356 GCACCTGCATGCTCTGCAAGTAACTAACACTCAAGGAACCAATGGCTTTGTTGGAGGTGGCTGT 297
QY 1865 ACTGAAACTCATTTGGCTGCATATATCAGACACAGACTCACAACGACCCAGAAAGCATT 1924
DB 296 ACTGAAACTCATTTGGCTGCATATATCAGACACAGACTCACAACGACCCAGAAAGCATT 237
QY 1925 CTCAAAGATGATGAATGTACTCAAAACAGAACTTCAATTAATTGCTGAAGCATTTTGCAGT 1984
DB 236 CTCAAAGATGATGAATGTACTCAAAACAGAACTTCAATTAATTGCTGAAGCATTTTGCAGT 177
QY 1985 GCCTAGAACTCTTTGCTGCTCTTTAGAACATGATGGAGTGGAATTCCTACTGACATG 2044
DB 176 GCCTAGAACTCTTTGCTGCTCTTTAGAACATGATGGAGTGGAATTCCTACTGACATG 117
QY 2045 AAGTATGGACACCTTTGGTTCAGTTTCAGGCAGATTCCTCCCTGTGCTGCTAACTGGCCAGAT 2104
DB 116 AAGTATGGACACCTTTGGTTCAGTTTCAGGCAGATTCCTCCCTGTGCTGCTAACTGGCCAGAT 57
QY 2105 TTGCTTTCAGTGTGGCTGTGATATACAATAGCCAGGAAGAACTCAACTGGTC 2160
DB 56 TTGCTTTCAGTGTGGCTGTGATATACAATAGCCAGGAAGAACTCAACTGGTC 1

RESULT 15
ABZ32791/C
ID ABZ32791 standard; cDNA; 421 BP.
XX
XX AC ABZ32791;
XX
XX DT 30-JAN-2003 (first entry)
XX
XX DE Human colon tumour related cDNA for CoSub-42 SEQ ID NO:147.
XX
XX KW Human; colon cancer; colon tumour; immunotherapy; diagnosis; cancer;
XX tumour; immune response; immunostimulant; cytostatic; vaccine; gene; ss.
XX
XX OS Homo sapiens.
XX
XX PN WQ200283070-A2.
XX
XX PD 24-OCT-2002.
XX
XX PF 09-APR-2002; 2002WO-US011475.
XX
XX PR 10-APR-2001; 2001US-00833263.
XX 03-AUG-2001; 2001US-00922217.
XX 19-DEC-2001; 2001US-00025380.
XX
XX (CORI-) CORIXA CORP.
XX
XX Xu J, Lodes MJ, Secretist H, Benson DR, Meagher MJ, Stolk JA;
PI Wang T, Jiang Y, Smith CL, King GE, Wang A, Clapper JD, Skeiky YAW;
PI Fanger GR, Vedvick TS, Carter D;
XX
XX WPI; 2003-067548/06.
XX
XX New polynucleotide, useful for the preparation of a composition for
PT stimulating an immune response against, or treating, cancer.
XX
XX Example 2; Page 176-177; 537pp; English.
XX
XX The present invention describes compounds (I) for the immunotherapy and
CC diagnosis of colon cancer. Also described: (1) a method for detecting the
CC presence of cancer in a patient; (2) a method for stimulating and/or
CC expanding T cells specific for a tumour protein; (3) an isolated T cell
CC population comprising T cells prepared by the method of (2); (4) a method
CC for stimulating an immune response in a patient; (5) a method for
CC treating cancer in a patient; and (6) a method for inhibiting the
CC development of cancer in a patient. (I) have immunostimulant and
CC cytostatic activities and can be used in vaccines. ABZ32646 to ABZ33725
CC and ABP55343 to ABP55391 represent human colon cancer/tumour related
CC sequences used in the exemplification of the present invention
XX
XX Sequence 421 BP; 113 A; 100 C; 91 G; 116 T; 0 U; 1 Other;

Query Match 16.7%; Score 413.4; DB 8; Length 421;
Best Local Similarity 99.5%; Pred. No. 4.7e-59;
Matches 414; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1745 TTGCTTCTCTGCAACAGAAATGACACTGCTGGATGAGCTGAAGCTCAGTGTCAAGC 1804
DB 416 TGGCTTCTCTGCAACAGAAATGACACTGCTGGATGAGCTGAAGCTCAGTGTCAAGC 357
QY 1805 GCACCTGCATGCTCTGCAAGTAACTAACACTCAAGGAACCAATGGCTTTGTTGGAGGTGGCTGT 1864
DB 356 GCACCTGCATGCTCTGCAAGTAACTAACACTCAAGGAACCAATGGCTTTGTTGGAGGTGGCTGT 297
QY 1865 ACTGAAACTCATTTGGCTGCATATATCAGACACAGACTCACAACGACCCAGAAAGCATT 1924
DB 296 ACTGAAACTCATTTGGCTGCATATATCAGACACAGACTCACAACGACCCAGAAAGCATT 237
QY 1925 CTCAAAGATGATGAATGTACTCAAAACAGAACTTCAATTAATTGCTGAAGCATTTTGCAGT 1984
DB 236 CTCAAAGATGATGAATGTACTCAAAACAGAACTTCAATTAATTGCTGAAGCATTTTGCAGT 177

Qy	1985	GCCTAGAACTCTGTTGTTGGCTCTTTAGAACATGATGGAGGTGAAATTTCTCACTGACATG	2044
Db	176		
Qy	2045	AAGTATGGACACCTTTGGTCAGTTCAGGCAGATTCTCCCTGTGTTGCTAACTGGCCAGAT	2104
Db	116		
Qy	2105	TTGCTTTTCACAGTGTGGCTGTGGATTATACAATAGCCAGGAAGAACTCAACTGGTC	2160
Db	56		

Search completed: September 6, 2005, 12:00:12
Job time : 1323 secs

THIS PAGE LEFT BLANK

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 3, 2005, 04:44:12 ; Search time 178 Seconds
(without alignments)
1639.804 Million cell updates/sec

Title: US-10-616-263-30

Perfect score: 2957

Sequence: 1 MSRLAKPKSLCKSEPLTTE.....VAVETANLIWLSYVIEDGN 570

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Uniprot_03.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2936	99.3	570	1 MKKS HUMAN	Q9npj1 homo sapien
2	2919	98.7	570	2 Q81222	Q81222 homo sapien
3	2904	98.2	570	2 Q9H6M6	Q9H6M6 homo sapien
4	2334	78.9	570	2 Q8BQ33	Q8BQ33 m mus-muscu
5	2331	78.8	570	1 MKKS MOUSE	Q9j170 mus musculu
6	2021	68.3	502	2 Q9DBF3	Q9dbf3 mus musculu
7	1185.5	40.1	563	2 Q7ZVV0	Q7zvv0 brachydanio
8	516	17.5	121	2 Q8K3C1	Q8k3c1 mus musculu
9	279	9.4	553	2 Q8ZTF8	Q8ztf8 pyrobaculum
10	270	9.1	557	2 Q59662	Q59662 pyrodictium
11	259.5	8.7	563	2 Q877H0	Q877h0 acidianus t
12	257.5	8.7	538	1 THSB METTH	O26885 methanobact
13	256.5	8.7	560	1 THSB SULSH	P46219 sulfolobus
14	255	8.6	490	1 THSA SULAC	Q9v2t5 sulfolobus
15	253.5	8.6	546	2 Q6L132	Q6l132 picropilus
16	253.5	8.6	559	1 THSA SULSO	Q9v289 sulfolobus
17	253	8.6	552	2 Q8TU11	Q8tu11 methanosarc
18	252.5	8.5	559	1 THSA SULTO	O24734 sulfolobus
19	248.5	8.4	554	1 THSA AERPE	Q9vdk6 aeropyrum p
20	246	8.3	551	2 Q8PX43	Q8px43 methanosarc
21	244	8.3	549	2 Q8ZVU7	Q8zvuv7 pyrobaculum
22	243	8.2	542	2 Q7R886	Q7r886 plasmodium
23	238	8.0	545	1 THSA ARCFU	O28045 archaeoglob
24	236	8.0	554	2 Q6BY12	Q6by12 debaryomyce
25	235	7.9	542	2 Q8PXK0	Q8pxk0 methanosarc
26	234.5	7.9	546	1 TCPO SCHPO	P78921 schizosach
27	234.5	7.9	561	2 Q6C9H3	Q6c9h3 yarrowia li
28	234.5	7.9	572	2 Q59663	O59663 pyrodictium
29	234	7.9	495	1 THSB SULAC	Q9v2t4 sulfolobus
30	233	7.9	528	2 Q8IDC7	Q8idc7 plasmodium
31	232	7.8	517	2 Q7RKG4	Q7rkg4 plasmodium

```

32 232 7.8 535 1 THSG SULSH
33 230 7.8 543 2 Q877G8
34 229.5 7.8 542 1 THS METJA
35 229 7.7 542 2 Q815C4
36 228.5 7.7 502 1 THSB DESMO
37 227.5 7.7 559 2 Q6CY04
38 227 7.7 543 2 Q8THU8
39 226 7.6 535 1 THSG SULSO
40 226 7.6 548 1 THSB_AERPE
41 226 7.6 558 2 Q751B1
42 225.5 7.6 545 1 THSA THEAC
43 225 7.6 544 1 TCFG HUMAN
44 225 7.6 563 2 Q6CRG0
45 224.5 7.6 547 2 Q751Y6

```

ALIGNMENTS

```

RESULT 1
MKKS HUMAN
ID MKKS_HUMAN STANDARD; PRT; 570 AA.
AC Q9NFU1;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE McKusick-Kaufman/Bardet-Biedl syndromes putative chaperonin.
GN Name=MKKS;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]_
RP SEQUENCE FROM N.A., VARIANTS MKKS CYS-37; TYR-84 AND SER-242, AND
RP VARIANTS VAL-49 AND CYS-517.
RX MEDLINE=20264373; PubMed=10802661; DOI=10.1038/75637;
RA Stone D.L., Slavotinek A.M., Bouffard G.G., Banerjee-Basu S.,
RA Baxevasis A.D., Barr M., Biesecker L.G.;
RT "Mutation of a gene encoding a putative chaperonin causes McKusick-
RT Kaufman syndrome.";
RL Nat. Genet. 25:79-82 (2000).
RN [2]
RP SEQUENCE FROM N.A.
TISSUE=amygdala;
RA Ottenwaelder B., Obermaier B., Meves H.-W., Weil B., Wiemann S.;
RA Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
MEDLINE=21638749; PubMed=11780052; DOI=10.1038/414865a;
RA Deloukas P., Matthews L.H., Aehurst J.L., Burton J., Gilbert J.G.R.,
RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,
RA Bailey J., Barlow K.P., Bates K.N., Beard L.M., Beare D.M.,
RA Beakley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.V., Clee C.M.,
RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
RA Coulson A., Coville G.J., Deadman R., Dhani P.D., Dunn M.,
RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
RA Graham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
RA Huckle B., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
RA Lehesvallo M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
RA Marsh V.L., Martin S.L., McConachie L.J., McLay K., McMurray A.A.,
RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
RA Rice C.M., Rose M.T., Scott C.E., Sehra H.K., Showkeen R., Sims S.,
RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
RA Swann R.M., Symamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
RA Tracey A., Tromamore A.C., Vaudin M., Wall M., Wallis J.M.,
RA Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,
RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,

```


Db 1 MSRLKPKSLCKSEPLTTERVTTLSVLKRIIVTSCYGPSGRLLKQLHNGFGYVCTTSQS 60
 QY 61 SALLSHLLVTHPIKILITASIQNHVSFSDCGLFTAILCCNLNENVRGLTPTTVIRLN 120
 Db 61 SALLSHLLVTHPIKILITASIQNHVSFSDCGLFTAILCCNLNENVRGLTPTTVIRLN 120
 QY 121 KHLLSLCISYKSDTCGCRIPVDFSSQTQLCLVRSILTSKPACMLTRKETEHSALILR 180
 Db 121 KHLLSLCISYKSDTCGCRIPVDFSSQTQLCLVRSILTSKPACMLTRKETEHSALILR 180
 QY 181 AFLTTIPENAEGHIIIGKSLIVPLKQGVDSITVPLGILTEMSVQLMRLPIKKSALK 240
 Db 181 AFLTTIPENAEGHIIIGKSLIVPLKQGVDSITVPLGILTEMSVQLMRLPIKKSALK 240
 QY 241 VALFCTTSLSGDTSDBGTGTVVSVGSLENAVLQDLNLRQLISDHVDLVLCQKVIHPS 300
 Db 241 VALFCTTSLSGDTSDBGTGTVVSVGSLENAVLQDLNLRQLISDHVDLVLCQKVIHPS 300
 QY 301 LKQFLNHRHIIAIDRIGVTLMPELTKMTGTQPIGSLGSCIPNSYGSVKDVCTAKFGSQHF 360
 Db 301 LKQFLNHRHIIAIDRIGVTLMPELTKMTGTQPIGSLGSCIPNSYGSVKDVCTAKFGSKHF 360
 QY 361 PHILPNEATICSLLCNRNDAWDELKLTCTQALHVLQTLKEPWALLGGGCTETHLAAY 420
 Db 361 PHILPNEATICSLLCNRNDAWDELKLTCTQALHVLQTLKEPWALLGGGCTETHLAAY 420
 QY 421 IRKTHNDPESILKDDCTQTELQIAEAFCSALESVVGSLEHDDGGGILTDMKYGHLSV 480
 Db 421 IRKTHNDPESILKDDCTQTELQIAEAFCSALESVVGSLEHDDGGGILTDMKYGHLSV 480
 QY 481 QADSPCVANPDLSSQCGGLYNSQBELNWSFLRSTRPFPVQSCLPHEAVGSASNLTLTD 540
 Db 481 QADSPCVANPDLSSQCGGLYNSQBELNWSFLRSTRPFPVQSCLPHEAVGSASNLTLTD 540
 QY 541 CLTAKLSGLQVAVETANLIWDLISYVIEDKN 570
 Db 541 CLTAKLSGLQVAVETANLIWDLISYVIEDKN 570

RESULT 2
 Q81222 ID Q81222 PRELIMINARY; PRT; 570 AA.
 AC Q81222;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE McKusick-Kaufman syndrome protein.
 GN Name=MKKS;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Skin;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Klausner R.D., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Udén T.B., Tohiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
 RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gumaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Vallalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettewam M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzywinski M.I., Skalska U., Smallos D.E., Schnerch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;

"Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences."; Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Skin;
 RA Strausberg R.; Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
 RL EMBL; BC028973; AAH28973.1; -.
 DR GO; GO:0005524; F-ATP binding; IEA.
 DR GO; GO:0051082; F:unfolded protein binding; IEA.
 DR GO; GO:0006457; P:protein folding; IEA.
 DR InterPro; IPR002423; Cpn60/TCP-1.
 DR InterPro; IPR008950; GroEL-ATPase.
 DR Pfam; PF00118; Cpn60 TCP1; 1.
 KW ATP-binding; Chaperone.
 SQ SEQUENCE 570 AA; 62330 MW; 01E57FF8AF7BA80 CRC64;

Query Match 98.7%; Score 2919; DB 2; Length 570;
 Best Local Similarity 99.1%; Pred. No. 7.3e-216;
 Matches 565; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSRLKPKSLCKSEPLTTERVTTLSVLKRIIVTSCYGPSGRLLKQLHNGFGYVCTTSQS 60
 Db 1 MSRLKPKSLCKSEPLTTERVTTLSVLKRIIVTSCYGPSGRLLKQLHNGFGYVCTTSQS 60
 QY 61 SALLSHLLVTHPIKILITASIQNHVSFSDCGLFTAILCCNLNENVRGLTPTTVIRLN 120
 Db 61 SALLSHLLVTHPIKILITASIQNHVSFSDCGLFTAILCCNLNENVRGLTPTTVIRLN 120
 QY 121 KHLLSLCISYKSDTCGCRIPVDFSSQTQLCLVRSILTSKPACMLTRKETEHSALILR 180
 Db 121 KHLLSLCISYKSDTCGCRIPVDFSSQTQLCLVRSILTSKPACMLTRKETEHSALILR 180
 QY 181 AFLTTIPENAEGHIIIGKSLIVPLKQGVDSITVPLGILTEMSVQLMRLPIKKSALK 240
 Db 181 AFLTTIPENAEGHIIIGKSLIVPLKQGVDSITVPLGILTEMSVQLMRLPIKKSALK 240
 QY 241 VALFCTTSLSGDTSDBGTGTVVSVGSLENAVLQDLNLRQLISDHVDLVLCQKVIHPS 300
 Db 241 VALFCTTSLSGDTSDBGTGTVVSVGSLENAVLQDLNLRQLISDHVDLVLCQKVIHPS 300
 QY 301 LKQFLNHRHIIAIDRIGVTLMPELTKMTGTQPIGSLGSCIPNSYGSVKDVCTAKFGSQHF 360
 Db 301 LKQFLNHRHIIAIDRIGVTLMPELTKMTGTQPIGSLGSCIPNSYGSVKDVCTAKFGSKHF 360
 QY 361 PHILPNEATICSLLCNRNDAWDELKLTCTQALHVLQTLKEPWALLGGGCTETHLAAY 420
 Db 361 PHILPNEATICSLLCNRNDAWDELKLTCTQALHVLQTLKEPWALLGGGCTETHLAAY 420
 QY 421 IRKTHNDPESILKDDCTQTELQIAEAFCSALESVVGSLEHDDGGGILTDMKYGHLSV 480
 Db 421 IRKTHNDPESILKDDCTQTELQIAEAFCSALESVVGSLEHDDGGGILTDMKYGHLSV 480
 QY 481 QADSPCVANPDLSSQCGGLYNSQBELNWSFLRSTRPFPVQSCLPHEAVGSASNLTLTD 540
 Db 481 QADSPCVANPDLSSQCGGLYNSQBELNWSFLRSTRPFPVQSCLPHEAVGSASNLTLTD 540
 QY 541 CLTAKLSGLQVAVETANLIWDLISYVIEDKN 570
 Db 541 CLTAKLSGLQVAVETANLIWDLISYVIEDKN 570

RESULT 3
 Q9H6M6 ID Q9H6M6 PRELIMINARY; PRT; 570 AA.
 AC Q9H6M6;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Hypothetical protein FLJ22088.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC	Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX	NCBI_TaxID=9606;
ON	[1]
OR	SEQUENCE FROM N.A.
RP	Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA	Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,
RA	Tanaka T., Nakamura Y., Isogai T., Sugano S.;
RL	Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RL	EMBL; AK025741; BAB15230.1; -
DR	GO; GO:0005524; F-ATP binding; IEA.
DR	GO; GO:001082; F-unfolded protein binding; IEA.
DR	GO; GO:0006457; P-protein folding; IEA.
DR	InterPro; IPR002423; Cpn60/TCP-1.
DR	InterPro; IPR008950; GroEL-ATPase.
DR	Pfam; PF00118; Cpn60_TCP1; 1.
KW	ATP-binding; Chaperone.
SQ	SEQUENCE 570 AA; 62347 MW; 2F3FE9459F472344 CRC64;
Query Match	
Best Local Similarity 98.2%; Score 2904; DB 2; Length 570;	
Matches 562; Conservative 2; Mismatches 6; Indels 0; Gaps 0;	
Qy	1 MSRLKAKPSLCKSEPLTTERVRTLVLKRVITSCYGPSGRLLKQLHNGFGVCTTSQ 60
Db	1 MSRLKAKPSLCKSEPLTTERVRTLVLKRVITSCYGPSGRLLKQLHNGFGVCTTSQ 60
Qy	61 SALLSHLLVTHPIKLITASTIQNVSSFDGCLFTAILCCNLINQVORGLTPTTVIRLN 120
Db	61 SALLSHLLVTHPIKLITASTIQNVSSFDGCLFTAILCCNLINQVORGLTPTTVIRLN 120
Qy	121 KHLLSLCISYLKSDTCGRIPVDFSSQIILCLVRSILTSKPACMLTRKTEHVSALILR 180
Db	121 KHLLSLCISYLKSDTCGRIPVDFSSQIILCLVRSILTSKPACMLTRKTEHVSALILR 180
Qy	181 AFLTTPENAEGHIIIGKSLIVPLKQGVDSITVLPGLIEMSEVQLMRLLPKKGSTALK 240
Db	181 AFLTTPENAEGHIIIGKSLIVPLKQGVDSITVLPGLIEMSEVQLMRLLPKKGSTALK 240
Qy	241 VALFCTTSLGSDTSDEGTVVYGVSVLENVAVLQQLNLGRQLISDHVDVLVLCQKVHPS 300
Db	241 VALFCTTSLGSDTSDEGTVVYGVSVLENVAVLQQLNLGRQLISDHVDVLVLCQKVHPS 300
Qy	301 LKQFLNMHRIIADIRIGVTLMPLTKMTGTQPIGSLGSI CPNSYGVSKVCTAKFGSQHF 360
Db	301 LKQFLNMHRIIADIRIGVTLMPLTKMTGTQPIGSLGSI CPNSYGVSKVCTAKFGSKHF 360
Qy	361 PHILPNEATICSLLICNRNDTAWDELKUTCTOTALHVQLTLTKPFWALLGGCGCTETHLAAY 420
Db	361 PHILPNEATICSLLICNRNDTAWDELKUTCTOTALHVQLTLTKPFWALLGGCGCTETHLAAY 420
Qy	421 IRKHTNDPESILKODECTQTELQIIAEAFCSALESVVGSLHDGGEILTDMKYGHLWSV 480
Db	421 IRKHTNDPESILKODECTQTELQIIAEAFCSALESVVGSLHDGGEILTDMKYGHLWSV 480
Qy	481 QADSPCVANWPDLLSCQCGGLYNSQEELNWSFLSTRPFPVQSCIPHEAVGSASNLITLD 540
Db	481 HADSPCVANWPDLLSCQCGGLYNSQEELNWSFLSTRPFPVQSCIPHEAVGSASNLITLD 540
Qy	541 CLTAKLSGLQVAVETANLILWLSYVIEDKN 570
Db	541 CLTAKLSGLQVAVETANLILWLSYVIEDKN 570
RESULT 4	
Q8BGQ3	PRELIMINARY; PRT; 570 AA.
AC	Q8BGQ3;
DT	01-MAR-2003 (TrEMBLrel. 23, Created)
DT	01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT	25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE	Mus musculus adult male olfactory brain cDNA, RIKEN full-length
DE	enriched library, clone:6430582D06 product:McKusick-Kaufman syndrome
DE	protein, full insert sequence (Mus musculus adult male olfactory brain

cDNA, RIKEN full-length enriched library, clone:6430593117
product:McKusick-Kaufman syndrome protein, full insert sequence).
DE Name=Mkks;
DN
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
RN
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
[2]
RN
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
[3]
RN
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;
RA The FANTOM Consortium,
RT "Analysis of the mouse transcriptome based on functional annotation of
60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
[4]
RN
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
[5]
RN
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
RA Sumi N., Iehii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
[6]
RN
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
RA Sumi N., Iehii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
[6]
RN
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
RA Sumi N., Iehii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
[6]
RN
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
RA Sumi N., Iehii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
[6]
RN
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
RA Sumi N., Iehii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
[6]
RN
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
RA Sumi N., Iehii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
[6]
RN
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
RA Sumi N., Iehii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
[6]
RN
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
RA Sumi N., Iehii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
[6]
RN
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
RA Sumi N., Iehii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
[6]
RN
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
RA Sumi N., Iehii Y., Nakamura S., Hazama M., Nishine T., Harada A.,

DE Mus musculus adult male liver cDNA, RIKEN full-length enriched
 DE library, clone:1300013E18 product:McKusick-Kaufman syndrome protein,
 DE full insert sequence.
 GN Name=Mkks;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=C57BL/6J; TISSUE=Liver;
 RC MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
 RX Carninci P., Hayashizaki Y.;
 RA "high-efficiency full-length cDNA cloning.";
 RT Meth. Enzymol. 303:19-44(1999).
 RL [2]
 RN SEQUENCE FROM N.A.
 RP STRAIN=C57BL/6J; TISSUE=Liver;
 RC MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
 RX RIKEN FANTOM Consortium;
 RA "Functional annotation of a full-length mouse cDNA collection.";
 RT Nature 409:685-690(2001).
 RL [3]
 RN SEQUENCE FROM N.A.
 RP STRAIN=C57BL/6J; TISSUE=Liver;
 RC The FANTOM Consortium;
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs.";
 RL Nature 420:563-573(2002).
 RN [4]
 RN SEQUENCE FROM N.A.
 RP STRAIN=C57BL/6J; TISSUE=Liver;
 RC MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
 RX Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
 RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
 RT "Normalization and subraction of cap-trapper-selected cDNAs to
 RT prepare full-length cDNA libraries for rapid discovery of new genes.";
 RL Genome Res. 10:1617-1630(2000).
 RN [5]
 RN SEQUENCE FROM N.A.
 RP STRAIN=C57BL/6J; TISSUE=Liver;
 RC MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
 RX Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P.,
 RA Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
 RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
 RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
 RT "RIKEN integrated sequence analysis (RISA) system-384-format
 RT sequencing pipeline with 384 multicapillary sequencer.";
 RL Genome Res. 10:1757-1771(2000).
 RN [6]
 RN SEQUENCE FROM N.A.
 RP STRAIN=C57BL/6J; TISSUE=Liver;
 RC Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,
 RA Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,
 RA Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,
 RA Inotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,
 RA Kawai J., Kojima Y., Konno H., Kouda M., Koyama S., Kurihara C.,
 RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,
 RA Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,
 RA Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,
 RA Sogabe Y., Suzuki H., Tagami M., Tegawa A., Takahashi F., Tanaka T.,
 RA Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
 RA Muramatsu M., Hayashizaki Y.;
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 RU EMBL; AK04967; BAB2722.1; -;
 DR MGD; MGI:1891836; Mkks.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0051082; F:unfolded protein binding; IEA.
 DR GO; GO:0006457; P:protein folding; IEA.

DR InterPro; IPR002423; Cpn60/TCP-1.
 DR InterPro; IPR008950; GroEL-ATPase.
 DR Pfam; PF00118; Cpn60_TCP1; 1.
 KW ATP-binding; Chaperone.
 SQ SEQUENCE 502 AA; 54611 MW; 6C4407F4BF19A6AD CRC64;
 Query Match 68.3%; Score 2021; DB 2; Length 502;
 Best Local Similarity 67.5%; Pred. No. 7.5e-147;
 Matches 385; Conservative 53; Mismatches 64; Indels 68; Gaps 1;
 QY 1 MSRLAKKPSLCKSEPLTATERRVTTLSVLKRVITVTCYGPSGRLLKQHNFGGVCVTTSS 60
 DB 1 MSRLAKKPSLCKTEPLTSEKVRSTLSVLKGVASCYGPSGRLLKQHNFGGVCVTTSS 60
 QY 61 SALLSHLVTHPIKILITASIQNHVSSPSCDGLFTAILCCNLIENVRIGLTPPTVIRLN 120
 DB 61 SALLRNLSVTHPVLKILTSSVQNVHSCFDCGLFTAILCCNLIENIRQLDLTPATAIKLN 120
 QY 121 KHLISICISYLSKSDTCGCRIPVDVFSQTLLCLVRSILTSKPCMLTRKETSHVSALILR 180
 DB 121 KYLLSUCTSYLSKSEACSCRIPVDVFSRTHFTLSVHSILTSKPCMLTRKETDHIGALILK 180
 QY 181 AFLITIPENAEGHIIILGKSLIVPLKQORVIDTVPGLIEMSEVQLMELLPIKKSTALK 240
 DB 181 AFLITIPESTEERMWLGKSIIVPLKQORVTDSTVPLGLLIEASEVQLRELLPTOK---- 235
 QY 241 VALFCTTSLSGDSTGEGTVVSYGVSLNAVLDQLNLGRQLISDHVDVLVLCQKVIHPS 300
 DB 236 ----- 235
 QY 301 LKQFLNMHRIIADRTGVTLMEPLTKMTGTQPIGSI GSCPNISVGSVDVCTAKFGSOHF 360
 DB 236 ----FFSERHVMADRVGVTLMESLSKVGTATPIGSLNPIVSTYGVKDVCSARFGSKRP 292
 QY 361 FHLIPNEATICSILLNCNRNDTAWDELKLTCTQALHVLQTLKEPWALLGGCGTETHLAAY 420
 DB 293 FHLIPNEAVCTLLLCNRNDTAWSELKLTCTQAMHVLQTLKEPWALLGGCGTETHLAAY 352
 QY 421 IRKHTNDPESILKODECTQTELOLIAEAFCSALESVGSLEHDDGGEILTMKYGHLSV 480
 DB 353 VRHKVHHEAEAIVRDDGCTQAKLHVAAEAFCSALESVAGSLEHDDGGEILIDTKYGHLSW 412
 QY 481 QADSPCVANWPDLLSQCGGLNYSQELNWSFLRSTRFPVPSQCLPHEAVGSASNLTD 540
 DB 413 QADSASVGNWSDTLSCGGLNYSQELNWSFLRSTRFPVPSQCLPHEAVGSASNLTD 472
 QY 541 CLTAKLSGLQVAVETANLIWLSYVIEDKN 570
 DB 473 CFTAKLSGLQVAVETANLIWLSYVIEDKN 502
 [1]
 RESULT 7
 Q7ZVV0 PRELIMINARY; PRT; 563 AA.
 AC Q7ZVV0;
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DE 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Similar to McKusick-Kaufman syndrome.
 GN ORFNames=zgc:55608;
 OS Brachydanio rerio (Zebrafish) (Danio rerio).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
 OC Cyprinidae; Danio.
 OX NCBI_TaxID=7955;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AB; TISSUE=Whole body;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,

Db 531 ---VLDSPFAKLNALNVAVEMASLVLDVKYIKDVN 563
 || :|||: | |||| :|: | :|:|:|:|
 RESULT 8
 Q8K3C1 PRELIMINARY; PRT; 121 AA.
 ID Q8K3C1
 AC Q8K3C1;
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 22-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Mkks protein.
 DE Name=Mkks;
 GN Mus musculus (Mouse).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CZECH II; TISSUE=Mammary tumor;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Uesdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards D.K., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.S., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettaman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzywinski M.I., Skalak U., Smallus D.E., Schnrcher A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CZECH II; TISSUE=Mammary tumor;
 RA Strausberg R.;
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC024359; A424359.1; -;
 DR MGD; MG1:1891836; Mkks.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0051082; F:unfolded protein binding; IEA.
 DR GO; GO:0006457; P:protein folding; IEA.
 DR InterPro; IPR008950; GROEL-ATPase.
 SQ SEQUENCE 121 AA; 13018 MW; 3110080E3CE2D0AF CRC64;
 Query Match 17.5%; Score 516; DB 2; Length 121;
 Best Local Similarity 80.2%; Pred. No. 6.8e-32;
 Matches 97; Conservative 13; Mismatches 11; Indels 0; Gaps 0
 Qy 1 MSRLKAKPSLCKSEPLTTERVTTLSVLKRVTCYCPGSRLLKQLHNGFGCYVTTSSQS 60
 Db 1 MSRLKAKPSLCKTEPLTSEKVRSTLSVLKGVFIASCYCPGSRLLKQLHNGLGGCYVTTSSQS 60
 Qy 61 SALLSHLVTHPTLKLTATSIQNHVSSFSDCGLFTALCCNLLENVORLGIPTPTVIRLN 120
 Db 61 SALLRNLSTVHPVLKVTLSVQNHVSFCSCGLFTALCCNLLENIRQLDLTPATAIKLN 120
 Qy 121 K 121
 Db 121 K 121
 RESULT 9
 Q8ZTF8


```

Db 21 EALRNNTLAQVLAELVSSLPGLDKMLVDSFGD-VTTNDGATILKEMEIQHPAAKL 79
Qy 77 LTASIQNHVSFSQGLFTAILCCNLJ-----ENVORGLPTTIVIRLNKHLISICISYLK 132
Db 80 MVEVAKAQDAEVDG-GTTSVAVLAGMLLDRAENLDENIHTPTIIEGYKALDYALGELS 138
Qy 133 SDTCGRIPVDFSTQILLCLVRSILTSK----PACMLTRKETEHVSALILRAPLTIPE 188
Db 139 K----IGVKVDNRQLLKRIASTSLYSKYVSGATW-----DKLTDMAVEAVLRVAEP 188
Qy 189 NAEG--HIIIGKSLIVPLKQGVDSVTLPGILLIEMSEVOLMRLPIPKKSTALKVALFCT 246
Db 189 RGDGTYSVRLDRVKIEKKGGSLDSQLVEGIVLDKEVH--PGMPKRVENAVIVLIDAP 246
Qy 247 TISGDTSDTCGTVVSVGSLEN---AVLDLNLNLCROLIS-----DHDV-- 289
Db 247 L-----EVEKPEITAKINITSPDQIKAFIDEEAKLLREWDRIYEIYAVMRKRDGMEPG 300
Qy 290 ----LVLCKQVHPSLKQFLNMHRIIAIDRIGVTLMPLTKMTGTQPIGSLGICPNSYG 345
Db 301 KAGIVWITQKIDVAGQHLAKKGIMAVRRVGRSLEKLEYATGGRIVSSLRDLKPEDLG 360
Qy 346 SVKDVCTAKGSOH--FFHLIPNEATICSLLCNENDTAWDELKLTCTQTAHVQLTLKE 403
Db 361 FAKLVEERKVGNDKQVFIEGCPNPKAV-TILLRGANDMVLDEAERSLNDALHVLNRVLR 419
Qy 404 PWALLGGCTETHLAAVIRKHTNDPESILKDDCTQTELQIAEAFCSALESVVGSLH 463
Db 420 PLIVPGGGAVEVELMRLR-----KYAESLGGKEQLAVEAYADALEEIPMWLA 468
Qy 464 DGG-----EILTDMKYGH 476
Db 469 SAGMDALQALMDLRLH 485

```

RESULT 11

```

ID Q877H0 PRELIMINARY; PRT; 563 AA.
AC Q877H0;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DE 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Chaperonin alpha subunit.
OS Acidianus tengchongensis.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Acidianus.
OX NCBI_TaxID=146920;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SS;
RA Ma O., Zhang Y.;
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY254173; AAP04526.1; -.
DR HSSP; P48424; IASS.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0051082; F:unfolded protein binding; IEA.
DR GO; GO:0006457; P:protein folding; IEA.
DR InterPro; IPR002194; Chaperonin_TCP-1.
DR InterPro; IPR001844; Chaperonin_Cpn60.
DR InterPro; IPR002423; Cpn60/TCP-1.
DR InterPro; IPR008950; GroEL-ATPase.
DR Pfam; PF00118; Cpn60 TCP1; 1.
DR PRINTS; PR00298; CHAPERONIN60.
DR PRINTS; PR00304; TCOMPLEXTCP1.
DR PROSITE; PS00750; TCP1_1; 1.
DR PROSITE; PS00751; TCP1_2; 1.
DR PROSITE; PS00995; TCP1_3; 1.
KW ATP-binding; Chaperone.
SQ SEQUENCE 563 AA; 60194 MW; 51022259612D8056 CRC64;

```

Query Match

Best Local Similarity 8.8%; Score 259.5; DB 2; Length 563;

Pred. No. 2.9e-11;

```

Matches 120; Conservative 89; Mismatches 218; Indels 83; Gaps 18;
Qy 29 LKEIVTSCVPSGRKLQKLNHGGVYCTTSSQSSALLSHLVTHPIKILKTASTQNHVSSF 88
Db 33 LAEMLRSSLPGLDKMLVDSFGD-VTTNDGATIVKEMEIQHPAAKLVEAAKAQDSEV 91
Qy 89 SDGGLFTAILCCNLJ-IENVQRL---GLTPTTVIR-----LNKHL-----LSLCISYLK 132
Db 92 GD-GTTSVAVLAGLPLEKAEASLVQNIHPTIIEGPKAFNKSLELLPOLATKVDYSDLN 150
Qy 133 SDTCGRIPVDFSTQILLCLVRSILTSKPAKMLTRKETEHVSALILRAPLTIPENAEG 192
Db 151 SAT-----ARDALKKIVVTMTSSK--FMAEGEELNKIMDIVIDAVTTTVAEPLPDG 198
Qy 193 --HIIIGKSLIVPLKQGVDSVTLPGILLIEMSEVOLMRLPIPKKSTALKVALFCTTSLG 250
Db 199 GYNVSLDLKIDKKGGTIEDSQLIRGIVLDKEVHAGMPRRVEKA---KIVLDASLEV 255
Qy 251 DTSDTGCGTVVSVGSLEN-----AVLDQ-----LNLGRQLTSDHVDLVLCQVIHPSL 301
Db 256 EKPE-----ISAKISITSPDQIKAFIDEEAKYLDKMDVKLASIGNVVICQKGGIDIA 308
Qy 302 KQFLNMHRIIAIDRIGVTLMPLTKMTGTQPIGSLGICPNSYGVSKVDCTAKGSOH-- 359
Db 309 QHFLAKRGILAVRRVGRSIEKLEKALGARIISIKDATPEDLGYAELVEERKVGNDKMW 368
Qy 360 FFHLIPNEATICSLLCNENDTAWDELKLTCTQTAHVQLTLKEPWALLGGCTETHLAA 419
Db 369 FIEGAKNPKAV-NILLRGANDMVLDEAERSINDALYSRLNLMPEYIVPGGGAIELELSA 427
Qy 420 YIRKHTNDPESILKDDCTQTELQIAEAFCSALESVVGSLHDDG---EILTDMKYG 475
Db 428 RLREYARS-----VGKKEQLAIEAYADALEEIPWLAETAGLAPISALMDLRAR 476
Qy 476 HLMSVQADSPCVANWPDLLSQCGCGLYNSQ 505
Db 477 HARG-----LTNGGVDVINGK 492

```

RESULT 12

```

ID THSB METH STANDARD; PRT; 538 AA.
AC O26885;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Thermosome beta subunit (Thermosome subunit 2) (Chaperonin beta subunit).
GN Name=thSB; OrderedLocusNames=MTH794;
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
OC Methanobacteriaceae; Methanothermobacter.
OX NCBI_TaxID=187420;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Delta H;
RX MEDLINE=98037514; PubMed=9371463;
RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jiواني N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA McDougal S., Shimer G., Goyal A., Pietrowski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
RT "Complete genome sequence of Methanobacterium thermoautotrophicum
RL J. Bacteriol. 179:7135-7155(1997).
CC -!- FUNCTION: Molecular chaperone; binds unfolded polypeptides in
CC vitro, and has a weak ATPase activity (By similarity).
CC -!- SUBUNIT: Forms a hetero-oligomeric complex of two stacked eight-
CC membered rings (By similarity).
CC -!- SIMILARITY: Belongs to the TCP-1 chaperonin family.
CC

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).

CC -----
 CC EMBL; AE000857; AAB85294.1; -;
 CC PIR; H69205; H69205.
 CC HSP; P48424; IAGD.
 CC InterPro; IPR002194; Chaperonin_TCP-1.
 CC InterPro; IPR001844; Chaprinin_Cpn60.
 CC InterPro; IPR002423; Cpn60/TCP-1.
 CC InterPro; IPR008950; GroEL-ATPase.
 CC Pfam; PF00118; Cpn60_TCPI_1.
 CC PRINTS; PR00298; CHAPERONIN60.
 CC PRINTS; PR00304; TCOMPLEXTCP1.
 CC PROSITE; PS00750; TCPI_1; 1.
 CC PROSITE; PS00751; TCPI_2; 1.
 CC PROSITE; PS00995; TCPI_3; 1.
 CC KW ATP-binding; Chaperone; Complete proteome.
 CC SQ SEQUENCE 538 AA; 57885 MW; 493CC28C552461E3 CRC64;

Query Match 8.7%; Score 257.5; DB 1; Length 538;
 Best Local Similarity 24.4%; Pred. No. 3.9e-11;
 Matches 115; Conservative 84; Mismatches 214; Indels 65; Gaps 19;

QY 17 LTTRVTRTLVSLKRIVTCYSGRGLKQHNFGVYCTTSQSALLSHLLVTHPIIKI 76
 DB |||||
 QY 33 ILAETVRTTL-----GPKGMDKMLVDSIGDIV-VTNGVILKEMDIEHPAAKM 80
 DB |||||
 QY 77 LTASIQNHVSSFSDCGLFTAILCNLI-----ENVQRIGLPTTVIRLNKHLISLCISYLK 132
 DB |||||
 QY 81 LVEVAKTQDEVDG-GTTTAVITAGELLKKAENLLEMEIHP-TIANGVRAQAEKAQEIL 138
 DB |||||
 QY 133 SDTCGCRIPVDFSTQILLCIVRSILTSKPCMLTRKETHVSAIILRAFLLTIPENAE- 191
 DB |||||
 QY 139 DD-----IAIDASDRDTLMKVAMTAMTGK-----TEKAREPLAEILVDA-VKQVEEDGEV 188
 DB |||||
 QY 192 --GHILGSLIVPLKQORVIDSTVLPGLIEMSEVQLMRLPIKSKTALKFCTTSL 249
 DB |||||
 QY 189 EKDHIEKK-----EGAAVDDSTLVQGVILDKERVH--PCMP-KKVENAKIALNCPI- 239
 DB |||||
 QY 250 GDTSDTGEFTVSVYGVSLNAVLDQLNLGRQLISDHVD-----LVLCQKVIHPSLKQFL 305
 DB |||||
 QY 240 -EVKETEVDABEIRITDPSQMAFTQEQQEIMRDWNSIVDTGANVLCQKGIIDLAQHYL 298
 DB |||||
 QY 306 NMHRIIADIGVTLMEPLTKMTGTQPIGSLGSCIPNSYGVKDVCTAKEP-GSQHPFHLI 364
 DB |||||
 QY 299 AKAGVLAVRRYKSDMEKLSKATGANIVTNIEDLSPEDLGEAGVWSEKKISGSEMIFFVEE 358
 DB |||||
 QY 365 PNEATICSLLLCNRNDTAWDELKLTCTALHVLQTLKEPWLALGGGCTETHLAAYIRHK 424
 DB |||||
 QY 359 CKPEKAVTILVRGSTEHVSEVERAIEDAIGVNAATVEDKGVAGGAPAEIETA----- 412
 DB |||||
 QY 425 THNDPESILKD--DECTQTQLQLTAFAFCSALESVWGSLEHDDG-----EILTDMKYGH 476
 DB |||||
 QY 413 -----KRLKDYADISGRE-QLAVSAFAEALEIVPKTLAENAGLSIDVLVDLRAAH 463
 DB |||||

RESULT 13

THSA SULSH STANDARD; PRT; 560 AA.
 ID THSA SULSH
 AC P46219;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE Thermosome alpha subunit (Thermosome subunit 1) (Chaperonin alpha subunit) (Thermophilic factor 55 alpha) (TF55-alpha) (Ring complex alpha subunit) (Thermophilic factor 56).
 DE Name=thaA; Synonyms=tf56;
 GN Sulfolobus shibatae.

OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
 OC Sulfolobus.
 OX NCBI_TaxID=2286;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN=ATCC 51178 / DSM 5389 / B12;
 RX MEDLINE=96074780; PubMed=7473746;
 RA Kagawa H.K., Osipiuk J., Maltsev N., Overbeek R., Quate-Randall E.,
 RA Joachimiak A., Trent J.D.;
 RT "The 60 kDa heat shock proteins in the hyperthermophilic archaeon
 RT Sulfolobus shibatae.";
 RL J. Mol. Biol. 253:712-725 (1995).
 CC -!- FUNCTION: Molecular chaperone; binds unfolded polypeptides in
 CC vitro, stimulates protein folding and has ATPase activity.
 CC -!- SUBUNIT: Forms a hetero-oligomeric complex of two stacked nine-
 CC membered rings; one of alpha and the other of beta subunits.
 CC -!- INDUCTION: By heat shock.
 CC -!- SIMILARITY: Belongs to the TCP-1 chaperonin family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).

CC -----
 CC EMBL; L34691; AAA87624.1; -;
 CC PIR; S59859; S59859.
 CC HSP; P48424; IASS.
 CC InterPro; IPR002194; Chaperonin_TCP-1.
 CC InterPro; IPR001844; Chaprinin_Cpn60.
 CC InterPro; IPR002423; Cpn60/TCP-1.
 CC InterPro; IPR008950; GroEL-ATPase.
 CC Pfam; PF00118; Cpn60_TCPI_1.
 CC PRINTS; PR00298; CHAPERONIN60.
 CC PRINTS; PR00304; TCOMPLEXTCP1.
 CC PROSITE; PS00750; TCPI_1; 1.
 CC PROSITE; PS00751; TCPI_2; 1.
 CC PROSITE; PS00995; TCPI_3; 1.
 CC KW ATP-binding; Chaperone; Direct protein sequencing; Heat shock.
 CC SQ SEQUENCE 560 AA; 59703 MW; F42C945E74565604 CRC64;

Query Match 8.7%; Score 256.5; DB 1; Length 560;
 Best Local Similarity 24.3%; Pred. No. 4.9e-11;
 Matches 115; Conservative 84; Mismatches 221; Indels 53; Gaps 16;

QY 29 LKRVITSCYGPSGRGLKQHNFGVYCTTSQSALLSHLLVTHPIIKILTASIQNHVSSP 88
 DB |||||
 QY 31 LAEMLRSSLGPKGLDKMLDSFGD-VTIINDGATIVKDMEIQHPPAKLLIVEAKAQDAEV 89
 DB |||||
 QY 89 SDGGLFTAILCCNLINQVRL---GLTPTTVIRLNKHLISLCISYLVKSTCGCRIPVDFS 145
 DB |||||
 QY 90 GDGTTSAVVLGALKEKESLLDQNTHTPTIIIEGYKAVTKALELPQ--LGTRIDIRDL 147
 DB |||||
 QY 146 STQILLCLVRSI-----LTSKPACMLTRKETHVSAIILRAFLLTPE---NAEGHILGK 198
 DB |||||
 QY 148 NSSVARDTLRKIAFTTTLASK--FIAEGAEINKIIDMVIDA-IVNVAEPLPNGVNVSLDL 204
 DB |||||
 QY 199 SLIVPLKQORVIDSTVLPGLIEMSEVQLMRLPIKSKTALKVALPCTTLTSGTSDTGE 258
 DB |||||
 QY 205 IKIDKKKGSIEDSVLVKGLVDKEVVH--PCMP-PRVTAKAVLDALEVEKPE---- 257
 DB |||||
 QY 259 TVVVSYGVSLEN-----AVLDQ---LLNLGRQLISDHVDLVLCQKVIHPSLKQFLNMHR 309
 DB |||||
 QY 258 ---ISAKIGITSPEQIKAFIDBESKYLKDMVDKIASIGANVVICQKGIIDIAQHFLLAKG 314
 DB |||||
 QY 310 ITAIDRIGVTLMEPLTKMTGTQPIGSLGSCIPNSYGVKDVCTAKEGSOH--FFHLIPNE 367
 DB |||||
 QY 315 ILAVRVKRSDETEKLEKALGARIISIKDPTDDLGLYAEILVEERRVGNQDMKVFIEGAKNL 374
 DB |||||
 QY 368 ATICSELLCNRNDTAWDELKLTCTALHVLQTLKEPWLALGGGCTETHLAAYIRKTHN 427
 DB |||||

Db 375 KAV-NILGRSNDMALDEAERSINDALHARNILLEPVILPGGGAIELEAMKLRAYS 433
Qy 428 DPESILKDDCTQTELOLIAEAFCSALESVVGSLHDGG---EILTDMKYGH 476
Db 434 -----VGKKEQLAEAFADALEEPTILAEATAGLEAISALMDLRARH 475

RESULT 14

THSA SULAC
ID THSA SULAC STANDARD; PRT; 490 AA.
AC Q9V2T5;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Thermosome alpha subunit (Thermosome subunit 1) (Chaperonin alpha subunit) (Thermophilic factor 55 alpha) (TF55-alpha) (Fragment)
GN Name=tsa;
OS Sulfolobus acidocaldarius.
OC Archaea; Euryarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=2285;
RN [1]
SEQUENCE FROM N.A.
RP MEDLINE=59439500; PubMed=10508614; DOI=10.1016/S0960-9822(99)80457-6;
RA Archibald J.M., Logsdon J.M. Jr., Doolittle W.F.;
RT "Recurrent paralogy in the evolution of archaeal chaperonins.";
RL Curr. Biol. 9:1053-1056(1999).
CC -1- FUNCTION: Molecular chaperone; binds unfolded polypeptides in vitro, and has a weak ATPase activity (By similarity).
CC -1- SUBUNIT: Forms a hetero-oligomeric complex of two stacked eight-membered rings (By similarity).
CC -1- SIMILARITY: Belongs to the TCP-1 chaperonin family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF149923; AAF03364.1; -.
DR HSSP; P48424; IASS.
DR InterPro; IPR002194; Chaperonin TCP-1.
DR InterPro; IPR001844; Chaprin Cpn60.
DR InterPro; IPR002423; Cpn60/TCP-1.
DR InterPro; IPR008950; GroEL-ATPase.
DR Pfam; PF00118; Cpn60 TCP1; 1.
DR PRINTS; PR00298; CHAPERONIN60.
DR PRINTS; PR00304; TCOMPLEXTCP1.
DR PROSITE; PS00750; TCP1_1; 1.
DR PROSITE; PS00751; TCP1_2; 1.
DR PROSITE; PS00995; TCP1_3; 1.
KW ATP-binding; Chaperone.
FT NON TER 1 1
FT NON TER 490 490
SQ SEQUENCE 490 AA; 52843 MW; 07C0A186600AE868 CRC64;

Query Match

Best Local Similarity 8.6%; Score 255; DB 1; Length 490;
Matches 107; Conservative 90; Mismatches 219; Indels 50; Gaps 14;

Qy 35 SCVGPSRLKQHNFGFYVCTTSQSALLSHLVTHPILKILTASTQNHVVSFSCGLF 94
Db 3 SSGLPGLDKMLDSFGD-VTIINDGATIVKMEIQHPAAKLIVAEAKQAQAEVGD-GTT 60
Qy 95 TAILCCNLI-----ENVORGLTPTVIRLNKHLISLCSYKSDTCGCRIPVDFSTQIL 150
Db 61 SAVVLACGLLDKAEELLEQNVHPITIIDGYKKALTKALEII--DQLSLKIDVNDLSPTA 118
Qy 151 LCLVRSILTSKPCACMLTKRTEHVSALI-----LRAFLTTPENAE--HIILGKSLIVPLK 205
Db 119 KAQLKKIVSTMSKLTAGAEIDKIIDLIAITVAEKRPDGTYNVPLDLIKDKKK 178

Qy 206 GORVIDSTVLPGLIEMSEVOLMRLPIKKTALKKVALFCTTSLSGTSDTGEGTVVVSYG 265
Db 179 GGSIESILVHGLVLDKEVHAGMPRVEKA---KIAVLDAALEVEKPE-----ISAK 228
Qy 266 VSLEN-----AVLDQ-----LLNLGRQLISDHVLDVLCQKVIHPSLKQFLNMHRIIADRI 316
Db 229 ISITSPEQIKSFLEDEARVLEKEMVDKLASIGANVVICQKIGIDVVAQHLAKKGLAVRV 288
Qy 317 GVTLMEPLTQWCTQTPIGSLGSCPNYSYGVKDVCTAKGSOH--PFHLIPNEATICSLL 374
Db 289 KRSDIEKLEKAGARIISIKDATPEDLGYAEIVERRIGNDKWVFEGAKNPRV-NIL 347
Qy 375 LCNRNDAWDELKLTCTQALHVLQTLKEPWALLGGCTETHLAAYTRHKTNDHPESILK 434
Db 348 LRGSNDMALDEAERSLNDALHSRNVLKPMIVAGGAVESELAALRLREYARS----- 400
Qy 435 DDECTQTEQLIABAFCSALESVVGSLHDGG---EILTDMKYGH 476
Db 401 ----VGKKEQLAEKFAALEEIPMILAEATAGMEPIQALMDLRARH 442

RESULT 15

Q6LI32
ID Q6LI32 PRELIMINARY; PRT; 546 AA.
AC Q6LI32;
DT 05-JUL-2004 (T-EMBLrel. 27, Created)
DT 05-JUL-2004 (T-EMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (T-EMBLrel. 27, Last annotation update)
DE Thermosome subunit.
GN OrderedLocusNames=PTO0735;
OS Picrophilus torridus.
OC Archaea; Euryarchaeota; Thermoplasmata; Thermoplasmatales;
OC Picrophilaceae; Picrophilus.
OX NCBI_TaxID=82076;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=DSM 9790 / ATCC 700027;
RC PubMed=15184674; DOI=10.1073/pnas.0401356101;
RA Fuetteler O., Angelov A., Liesegang H., Gottschalk G., Schleper C.,
RA Schepers B., Dock C., Antranikian G., Liebl W.;
RT "Genome sequence of Picrophilus torridus and its implications for life around pH 0.";
RL Proc. Natl. Acad. Sci. U.S.A. 101:9091-9096(2004).
DR EMBL; AE017261; AAT43320.1; -.
DR HSSP; P48424; IASS.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0051082; F:unfolded protein binding; IEA.
DR GO; GO:006457; P:protein folding; IEA.
DR InterPro; IPR002194; Chaperonin TCP-1.
DR InterPro; IPR001844; Chaprin Cpn60.
DR InterPro; IPR002423; Cpn60/TCP-1.
DR Pfam; PF00118; Cpn60 TCP1; 1.
DR PRINTS; PR00298; CHAPERONIN60.
DR PRINTS; PR00304; TCOMPLEXTCP1.
DR PROSITE; PS00750; TCP1_1; 1.
DR PROSITE; PS00751; TCP1_2; 1.
DR PROSITE; PS00995; TCP1_3; 1.
KW ATP-binding; Chaperone; Complete proteome.
SQ SEQUENCE 546 AA; 58820 MW; 864ED7D4CAF736A0 CRC64;

Query Match

Best Local Similarity 8.6%; Score 253.5; DB 2; Length 546;
Matches 129; Conservative 95; Mismatches 243; Indels 103; Gaps 21;

Qy 20 ERVTRTLVKLRIVTSCYGPSRLKQHNFGFYVCTTSQSALLSHLVTHPILKILTA 79
Db 36 DAVRTTL-----GPKGMDKMLVDSIGDIV-ITNDGATILKEMIDIDHTAKLVE 83
Qy 80 STONHVVSFSDCGLFTAILCCNLIENVQL---GLPTTVIRLNKHLISLCSYKSDTC 136
Db 84 ASKSQDTAVGDGTTTWWVLAGELLKQAESLLEQGV-HSTVIASGYH---LAVTEAKQLD 139

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 3, 2005, 04:43:28 ; Search time 170 Seconds
(without alignments)

1296.785 Million cell updates/sec

Title: US-10-616-263-30

Perfect score: 2957

Sequence: 1 MSRLKAKPSLCKSEPLTTE.....VAVETANLIWLSYVIEDKN 570

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_16Dec04.*

1: Geneseqp1980s.*

2: Geneseqp1990s.*

3: Geneseqp2000s.*

4: Geneseqp2001s.*

5: Geneseqp2002s.*

6: Geneseqp2003as.*

7: Geneseqp2003bs.*

8: Geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2950	99.8	570	3 AAY94912	Aay94912 Human sec
2	2936	99.3	570	3 AAY91944	Aay91944 Human cha
3	2936	99.3	572	5 ABP41775	Abp41775 Human ova
4	2904	98.2	570	7 ADJ69265	Adj69265 Human hea
5	2503	84.6	498	8 ADP43680	Adp43680 Human PMM
6	2325	78.6	474	4 ABG19636	Abg19636 Novel hum
7	2086	70.5	406	3 AAB43205	Aab43205 Human ORF
8	1092	36.9	222	3 AAB56236	Aab56236 Human sec
9	1076	36.4	223	3 AAB56218	Aab56218 Human sec
10	841	28.4	166	3 AAB56356	Aab56356 Human sec
11	432	14.6	86	3 AAG02893	Ag02893 Human sec
12	257.5	8.7	538	8 ADN18515	Adn18515 Bacterial
13	256.5	8.7	560	7 AAO24120	Aao24120 S shibata
14	248.5	8.4	554	7 AAO24111	Aao24111 A pernix
15	248.5	8.4	557	8 ADS44430	Ads44430 Bacterial
16	245	8.3	520	7 ADB70179	Adb70179 C. neofor
17	238.5	8.1	558	6 ABG74939	Abg74939 A. Gossyp
18	238	8.0	545	8 ADS42984	Ads42984 Bacterial
19	236	8.0	540	8 ADS22203	Ads22203 Bacterial
20	235	7.9	537	8 ADN20780	Adn20780 Bacterial
21	235	7.9	542	7 AAO24112	Aao24112 M mazi a
22	232	7.8	535	7 AAO24121	Aao24121 S shibata
23	229.5	7.8	542	8 ADN18378	Adn18378 Bacterial
24	227	7.7	559	4 AAM79916	Aam79916 Human pro
25	227	7.7	567	6 AABJ25946	Abj25946 Aspergill

ALIGNMENTS

RESULT 1

AAAY94912

ID AAY94912 standard; protein; 570 AA.

AC AAY94912;

XX 16-JUN-2000 (first entry)

DE Human secreted protein clone pp314_19 protein sequence SEQ ID NO:30.

XX Human; secreted protein; immunestimulant; immunosuppressant; virucide;
KW antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;
KW antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;
KW antichyroid; immune deficiency; severe combined immunodeficiency; SCID;
KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
KW connective tissue disease; multiple sclerosis; erythematosis;
KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
KW insulin dependent diabetes mellitus; graft-versus-host-disease;
KW autoimmune inflammatory eye disease; allergy.

XX Homo sapiens.

OS WO200009552-A1.

XX 24-FEB-2000.

PF 13-AUG-1999; 99WO-US018298.

PR 14-AUG-1998; 98US-0096622P.

PR 17-AUG-1998; 98US-0096815P.

PR 04-SEP-1998; 98US-0099229P.

PR 23-OCT-1998; 98US-0105368P.

PR 08-JAN-1999; 99US-0115234P.

PR 12-FEB-1999; 99US-0119931P.

PR 18-FEB-1999; 99US-0120575P.

PR 30-APR-1999; 99US-0132020P.

PR 11-AUG-1999; 99US-0148424P.

XX (GEMY) GENETICS INST INC.

PA Jacobs K, Mccoy JM, Lavallie ER, Collins-Racie LA, Evans C;

PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;

PI Wong GG, Clark HF, Fechtel K;

DR WPI; 2000-205979/18.

XX New polynucleotides encoding secreted proteins, which may have e.g.

PT

PT nutritional, chemokine, immune stimulating or suppressing, hematopoiesis
PT regulating, tissue growth, activin/inhibin antiinflammatory or tumor
PT inhibition activity.
XX

XX Claim 39; Page 499-501; 641pp; English.

CC AAA16618 to AAA16697 encode the human secreted proteins given in AAY94898
CC to AAY94980, isolated from human adult brain, adult thyroid, adult
CC retina, foetal carcinoma, adult blood, adult neural, foetal kidney, adult
CC placenta, adult testis, whole embryo, adult cartilage, kidney, foetal
CC brain, adult thymus, foetal placenta, adult uterus, adult tumour, and
CC adult bladder, cDNA libraries. The polynucleotides and proteins are
CC predicted to have biological activities which would make them suitable
CC for treating, preventing or ameliorating medical conditions in humans and
CC animals. The polynucleotides can be used as markers for tissues in which
CC the protein is preferentially expressed, as molecular weight markers on
CC Southern gels, and as chromosome markers or tags to identify chromosomes
CC or to map gene positions. The proteins can be used in the treatment of
CC immune deficiencies and disorders, such as severe combined
CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other
CC infections. These infections include human immunodeficiency virus (HIV),
CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
CC candidiasis. The proteins can be used to treat autoimmune disorders such
CC as connective tissue disease, multiple sclerosis, systemic lupus
CC erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation,
CC Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent
CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
CC autoimmune inflammatory eye disease. The proteins can also be used to
CC treat allergic conditions, such as asthma. AAA16698 to AAA16774 represent
CC probes for the human secreted proteins from the present invention
XX
SQ Sequence 570 AA;

Query Match 99.8%; Score 2950; DB 3; Length 570;
Best Local Similarity 99.8%; Pred. No. 2.7e-297;
Matches 569; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MSRLAKKPSLCSEPLTTERVETTLVLKRIYVTCYSGPSGRUKQIHNGFGVYCVTSOS 60
DB 1 MSRLAKKPSLCSEPLTTERVETTLVLKRIYVTCYSGPSGRUKQIHNGFGVYCVTSOS 60
QY 61 SALLSHLLVTHPIKILITASIQNHVSSFDCGLFTAILCCNLJENVQRLGLTPTTVIRLN 120
DB 61 SALLSHLLVTHPIKILITASIQNHVSSFDCGLFTAILCCNLJENVQRLGLTPTTVIRLN 120
QY 121 KHLISLCISYLSKSDTCGRIPVDFSSQIILLCLVRSILTSKPACMLTRKETEHVSALILR 180
DB 121 KHLISLCISYLSKSDTCGRIPVDFSSQIILLCLVRSILTSKPACMLTRKETEHVSALILR 180
QY 181 AFLTTPENAEGHIILGKSLIVPLKQORVIDSTVLPGLIEMSEVQIMRLPLPKKSTALK 240
DB 181 AFLTTPENAEGHIILGKSLIVPLKQORVIDSTVLPGLIEMSEVQIMRLPLPKKSTALK 240
QY 241 VALFCTTSLGSDTSGEVTWVSYGVSLENAVLDQLNLGRQLISDHVDLVLCQKVIHPS 300
DB 241 VALFCTTSLGSDTSGEVTWVSYGVSLENAVLDQLNLGRQLISDHVDLVLCQKVIHPS 300
QY 301 LKQFLNMHRIIAIDRIGVTLMELPTKMTGTQPIGSLGSIICPNISYGVKDVCTAKPGSQHF 360
DB 301 LKQFLNMHRIIAIDRIGVTLMELPTKMTGTQPIGSLGSIICPNISYGVKDVCTAKPGSQHF 360
QY 361 FHLIPNEATCSLLLCNRNDTANDELKLTCTQTAHVLTQLTKPEWALLGGCGTETHLAAY 420
DB 361 FHLIPNEATCSLLLCNRNDTANDELKLTCTQTAHVLTQLTKPEWALLGGCGTETHLAAY 420
QY 421 IRKTHNDPESILKDBDCTQTELQIAEAFCSALESVVGSLHDGGEILLTDMKYGHLWSV 480
DB 421 IRKTHNDPESILKDBDCTQTELQIAEAFCSALESVVGSLHDGGEILLTDMKYGHLWSV 480
QY 481 QADSPCVANWPDLLSQCGCGLYNSQBELNWSFLRSTRPPVPOSLPHEAVGASNLITLD 540
DB 481 QADSPCVANWPDLLSQCGCGLYNSQBELNWSFLRSTRPPVPOSLPHEAVGASNLITLD 540

QY 541 CLTAKLSGLQAVETANLIWLSYVIEDKN 570
DB 541 CLTAKLSGLQAVETANLIWLSYVIEDKN 570
RESULT 2
AAY91944
ID AAY91944 standard; protein; 570 AA.
XX
AC AAY91944;
XX
DT 19-JUL-2000 (first entry)
XX
DE Human chaperone protein 5 (HCHP-5).
XX
KW Human chaperone protein; HCHP-5; neurodegenerative; cancer; metabolic;
KW developmental; autoimmune; inflammatory; cell proliferation.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Modified-site 2
FT Modified-site 10 /note= "potential phosphorylation site"
FT Modified-site 19 /note= "potential phosphorylation site"
FT Modified-site 40 /note= "potential phosphorylation site"
FT Modified-site 87 /note= "potential phosphorylation site"
FT Modified-site 159 /note= "potential phosphorylation site"
FT Modified-site 167 /note= "potential phosphorylation site"
FT Modified-site 171 /note= "potential phosphorylation site"
FT Modified-site 185 /note= "potential phosphorylation site"
FT Binding-site 192.199 /note= "potential phosphorylation site"
FT /label= ATP/GTP-binding_site_motif_A
FT /note= "P loop"
FT Modified-site 237 /note= "potential phosphorylation site"
FT Modified-site 300 /note= "potential phosphorylation site"
FT Modified-site 319 /note= "potential phosphorylation site"
FT Modified-site 346 /note= "potential phosphorylation site"
FT Modified-site 352 /note= "potential phosphorylation site"
FT Modified-site 379 /note= "potential phosphorylation site"
FT Modified-site 381 /note= "potential glycosylation site"
FT Modified-site 400 /note= "potential phosphorylation site"
FT Modified-site 425 /note= "potential phosphorylation site"
FT Modified-site 439 /note= "potential phosphorylation site"
FT Modified-site 452 /note= "potential phosphorylation site"
FT Modified-site 504 /note= "potential phosphorylation site"
FT Modified-site 509 /note= "potential glycosylation site"
FT Modified-site 515 /note= "potential phosphorylation site"
FT Modified-site 516 /note= "potential phosphorylation site"
FT Modified-site 536 /note= "potential phosphorylation site"

FT	Modified-site	/note= "potential glycosylation site"	
FT	543		
FT	/note= "potential phosphorylation site"		
XX	W0200017358-A2.		
XX	30-MAR-2000.		
XX	22-SEP-1999;	99WO-US022027.	
XX	22-SEP-1998;	98US-00158642.	
PR	22-SEP-1998;	98US-0172221P.	
PR	19-JAN-1999;	99US-00233291.	
PR	19-APR-1999;	99US-00294698.	
PR	19-APR-1999;	99US-0172232P.	
XX	(INCY-) INCYTE PHARM INC.		
PA	Tang YT, Hillman JL, Yue H, Patterson C, Baughn MR, Batra S;		
XX	WPI; 2000-283583/24.		
DR	N-PSDB; AAA08567.		
XX	New purified polypeptides and polynucleotides encoding human chaperone		
FT	proteins, useful for diagnosing, treating and preventing disorders		
PT	associated with the expression human chaperone proteins.		
XX	Claim 1; Page 73-75; 88pp; English.		
PS	AA91940-15 are human chaperone proteins 1-6 (HCHP-1 to HCHP-6)		
CC	respectively. The sequences can be used to treat and prevent disorders		
CC	associated with altered expression or activity of HCHP comprising		
CC	administering a composition comprising the polypeptide or an antagonist		
CC	to a patient (claimed). The human chaperone proteins are also useful for		
CC	the diagnosis, treatment or prevention of neurodegenerative, metabolic,		
CC	developmental, autoimmune/inflammatory disorders and cell proliferative		
CC	disorders including cancer		
XX	Sequence 570 AA;		
XX	Query Match	99.3%; Score 2936; DB 3; Length 570;	
XX	Best Local Similarity	99.5%; Pred. No. 7.8e-296;	
XX	Matches 567; Conservative	2; Mismatches 1; Indels 0; Gaps 0;	
QY	1	MSRLEAKKSLCKSEPLTTERVITLSVLKRVITSCVGPGRKQLHNGFGGVCTTSQS	60
Db	1	MSRLEAKKSLCKSEPLTTERVITLSVLKRVITSCVGPGRKQLHNGFGGVCTTSQS	60
QY	61	SALLSHLLVTHPIKILITASIQNHVSSFSDCGLFTAILCCNLNENQRLGLTPTTVIRLN	120
Db	61	SALLSHLLVTHPIKILITASIQNHVSSFSDCGLFTAILCCNLNENQRLGLTPTTVIRLN	120
QY	121	KHLISCISVLKSDTCGRIPVDFSSQIILLVRSILTSKPCMLTRKETEHSALILR	180
Db	121	KHLISCISVLKSDTCGRIPVDFSSQIILLVRSILTSKPCMLTRKETEHSALILR	180
QY	181	AFLLTTPENAEGLIIGKSLIVPLKQGVDSITVLPGLIEMSEVOLMRLLPKKSTALK	240
Db	181	AFLLTTPENAEGLIIGKSLIVPLKQGVDSITVLPGLIEMSEVOLMRLLPKKSTALK	240
QY	241	VALFCTTSLGSDTSGEGTVVSVGYSLNENAVLDQLNLNLRQLISDHVDLVLCQVIHPS	300
Db	241	VALFCTTSLGSDTSGEGTVVSVGYSLNENAVLDQLNLNLRQLISDHVDLVLCQVIHPS	300
QY	301	LKQFLNHRHIIADRIQVTLMELPTKMTGTQPIGSGISCPNSYGVKDVCTAKFGSQHF	360
Db	301	LKQFLNHRHIIADRIQVTLMELPTKMTGTQPIGSGISCPNSYGVKDVCTAKFGSKHF	360
QY	361	FHLIPNEATCSLLCNRNNDTANDELKLTCTQTAHVLTQLTKEPWALLGGGCTETHLAY	420
Db	361	FHLIPNEATCSLLCNRNNDTANDELKLTCTQTAHVLTQLTKEPWALLGGGCTETHLAY	420
QY	421	IRKHTNDPESILKDDCTQTELQIAEAFCSALESVVGSLHDGGEILLTDMKYGHLWSV	480

Db	421	IRKHTNDPESILKDDCTQTELQIAEAFCSALESVVGSLHDGGEILLTDMKYGHLWSV	480
QY	481	QADSPCVANWPDLLSQCGCLYNQBELNWSFLSTRRPFVQSCLPHEAVGSASNLTLTD	540
Db	481	QADSPCVANWPDLLSQCGCLYNQBELNWSFLSTRRPFVQSCLPHEAVGSASNLTLTD	540
QY	541	CLTAKLSGLQAVAVETANLIWDLISYVIEDKN	570
Db	541	CLTAKLSGLQAVAVETANLIWDLISYVIEDKN	570
RESULT 3			
ABP411775			
ID	ABP41775	standard; protein; 572 AA.	
XX	AC	ABP41775;	
XX	XX		
DT	22-AUG-2002	(first entry)	
XX	XX		
DE	Human ovarian antigen HMAEL73, SEQ ID NO:2907.		
XX	XX		
KW	Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;		
KW	ovarian cancer; breast cancer; tumour; reproductive system disorder;		
KW	infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;		
KW	PCOS; ovarian cyst; dysmenorrhea; endocrine disorder; infection;		
KW	inflammatory condition; immune disorder; blood disorder;		
KW	cardiovascular disorder; respiratory disorder; neurological disorder;		
KW	gastrointestinal disorder; urinary system disorder; drug screening;		
KW	gene therapy; chromosome mapping; forensic analysis;		
KW	antibody preparation; cytostatic; immunomodulatory; neuroprotective;		
KW	antiinflammatory; gynaecological; reproductive.		
XX	XX		
OS	Homo sapiens.		
XX	XX		
PN	WO200200677-A1.		
XX	XX		
PD	03-JAN-2002.		
XX	XX		
PF	07-JUN-2001;	2001WO-US018569.	
XX	XX		
PR	07-JUN-2000;	2000US-0209467P.	
XX	XX		
PA	(HUMA-) HUMAN GENOME SCI INC.		
XX	XX		
PI	Birse CE, Rosen CA;		
XX	XX		
DR	WPI; 2002-147878/19.		
XX	N-PSDB; ABQ54852.		
PT	Isolated nucleic acid molecules encoding novel ovarian polypeptides,		
PT	useful in the prevention, treatment and diagnosis of cancer (e.g. ovarian		
PT	cancer), immune disorders, cardiovascular disorders and neurological		
XX	diseases.		
PS	Claim 11; SEQ ID NO 2907; 2922pp; English.		
XX	XX		
CC	The invention relates to 2175 novel human ovarian antigens (ABP41054-		
CC	ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also		
CC	encompasses polypeptides 90% identical and polynucleotides 95% identical		
CC	to the sequences of the invention. The invention additionally relates to		
CC	recombinant vectors and host cells comprising human ovarian antigen		
CC	polynucleotides, antibodies against human ovarian antigens, and the use		
CC	of ovarian antigen polynucleotides and polypeptides in diagnosing,		
CC	treating, prognosing or preventing various ovary and/or breast-related		
CC	disorders. Such conditions include ovarian cancer and breast cancer, and		
CC	metastatic tumours of ovarian or breast origin, reproductive system		
CC	disorders (e.g., infertility, disorders of pregnancy, anovulation,		
CC	polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine		
CC	disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic		
CC	shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and		
CC	vaginitis), immune disorders (e.g., congenital and acquired		
CC	immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),		

CC blood-related disorders (e.g., anaemia), cardiovascular disorders,
CC respiratory disorders, neurological disorders, gastrointestinal disorders
CC and urinary system disorders. Ovarian antigen polypeptides and
CC polynucleotides may also be used in screening for compounds which
CC modulate ovarian antigen expression or activity. The polynucleotides may
CC further be used for gene therapy, chromosome mapping, in the
CC identification of individuals and in forensic analysis, and the
CC polypeptides may be used as food additives or to prepare antibodies
CC useful in disease diagnosis, drug targeting and phenotyping. The present
CC sequence represents a human ovarian antigen of the invention. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pot_sequences
XX
SQ Sequence 572 AA;

Query Match 99.3%; Score 2936; DB 5; Length 572;
Best Local Similarity 99.5%; Pred. No. 7.8e-296;
Matches 567; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 MSRLKAKPSLCKSEPLTTERVTTLSVLKRVITSCYGPSGRKLQHLNFGGVCVCTTSQS 60
DB 3 MSRLKAKPSLCKSEPLTTERVTTLSVLKRVITSCYGPSGRKLQHLNFGGVCVCTTSQS 62
QY 61 SALLSHLLVTHPILKILTASIQNHVSSFSDCGLFTAILCCNLIENTVQRLGTPPTVIRLN 120
DB 63 SALLSHLLVTHPILKILTASIQNHVSSFSDCGLFTAILCCNLIENTVQRLGTPPTVIRLN 122
QY 121 KHLLSLCISYLKSDTCGCRIPVDFSSQIILLCLVRSILTSKPACMLTRKETEHSALILR 180
DB 123 KHLLSLCISYLKSDTCGCRIPVDFSSQIILLCLVRSILTSKPACMLTRKETEHSALILR 182
QY 181 AFLTTPENAEAGHIIIGKSLIPLKQORVIDSTVLPGLIEMSEVOLMRLPIKSKTALK 240
DB 183 AFLTTPENAEAGHIIIGKSLIPLKQORVIDSTVLPGLIEMSEVOLMRLPIKSKTALK 242
QY 241 VALFCTTSLGSDTSGEGTVVSYGVSLNVLNLDQLNLGRQLISDHVDLVLCQKVIHPS 300
DB 243 VALFCTTSLGSDTSGEGTVVSYGVSLNVLNLDQLNLGRQLISDHVDLVLCQKVIHPS 302
QY 301 LKQFLNMRHIIADRIQVTLMEPLTKMTGTQPIGSLGSIICPNISYGVSKDVCTAKFGSQHF 360
DB 303 LKQFLNMRHIIADRIQVTLMEPLTKMTGTQPIGSLGSIICPNISYGVSKDVCTAKFGSKHF 362
QY 361 FLHIPNEATCSLLLCNRNDTANDELKLTQCTALHVLQTLKPEFWALLGGCGCTETHLAAY 420
DB 363 FLHIPNEATCSLLLCNRNDTANDELKLTQCTALHVLQTLKPEFWALLGGCGCTETHLAAY 422
QY 421 IRKHTNDPESILKDDCTCTELQIAEAFCSALESVVGSLHDGGEILTDMKYGHLSV 480
DB 423 IRKHTNDPESILKDDCTCTELQIAEAFCSALESVVGSLHDGGEILTDMKYGHLSV 482
QY 481 QADSPCVANWPDLLSQCGCLYNSQBELNWSFLRSTRRPFVPOSCLPHEAVGASNLITLD 540
DB 483 QADSPCVANWPDLLSQCGCLYNSQBELNWSFLRSTRRPFVPOSCLPHEAVGASNLITLD 542
QY 541 CLTAKLSGLQVAVETANLIWDSYVIEDKN 570
DB 543 CLTAKLSGLQVAVETANLIWDSYVIEDKN 572

RESULT 4
ADJ69265
ID ADJ69265 standard; protein; 570 AA.
XX
AC ADJ69265;
XX
XX
DT 06-MAY-2004 (first entry)
XX
DE Human heat mitochondrial protein as a therapeutic target SeqID1071.
XX
KW mitochondrial; human; screening assay; diabetes mellitus;
KW Huntington's disease; osteoarthritis;

KW Leber's hereditary optic neuropathy; LHON;
KW mitochondrial encephalopathy lactic acidosis and stroke; MELAS;
KW myoclonic epilepsy ragged red fibre syndrome; MERRF; cancer;
KW neuroprotective; nontropic; antidiabetic; anticonvulsant; antiarthritic;
KW osteopathic; ophthalmological; cytostatic.
XX
OS Homo sapiens.
XX
XX WO2003087769-A2.
XX 23-OCT-2003.
XX
XX 04-APR-2003; 2003WO-US010870.
XX
XX 12-APR-2002; 2002US-0372843P.
XX 17-JUN-2002; 2002US-0389987P.
XX 20-SEP-2002; 2002US-0412418P.
XX
XX (MITO-) MITOKOR.
XX (BUCK-) BUCK INST AGE RES.
XX
XX Ghosh SS, Fahy ED, Zhang B, Gibson BW, Taylor SW, Glenn GM;
PI Warnock DE;
XX
XX WPI; 2003-845369/78.
DR
XX
XX Identifying a mitochondrial target for drug screening assays and for
PT treating diseases associated with altered mitochondrial function,
PT comprises detecting a modified polypeptide in a sample and correlating
PT with the disease.
XX
PS Claim 1; SEQ ID NO 1071; 180pp; English.
XX
CC This invention relates to novel mitochondrial targets that can be used
CC for therapeutic intervention in treating a disease associated with
CC altered mitochondrial function. Specifically, it refers to a method for
CC identifying proteins of the human heart mitochondrial proteome that are
CC useful for drug screening assays, as well as therapeutic targets. The
CC present invention describes a method for identifying such proteins that
CC can be used in the treatment of various diseases associated with altered
CC mitochondrial function including diabetes mellitus, Huntington's disease,
CC osteoarthritis, Leber's hereditary optic neuropathy (LHON), mitochondrial
CC encephalopathy lactic acidosis and stroke (MELAS), myoclonic epilepsy
CC ragged red fibre syndrome (MERRF) or cancer. Accordingly, these
CC compositions have neuroprotective, nontropic, antidiabetic,
CC anticonvulsant, antiarthritic, osteopathic, ophthalmological and
CC cytostatic activities. This polypeptide sequence is a human heart
CC mitochondrial protein of the invention.
XX
SQ Sequence 570 AA;

Query Match 98.2%; Score 2904; DB 7; Length 570;
Best Local Similarity 98.6%; Pred. No. 1.7e-292;
Matches 562; Conservative 2; Mismatches 6; Indels 0; Gaps 0;
QY 1 MSRLKAKPSLCKSEPLTTERVTTLSVLKRVITSCYGPSGRKLQHLNFGGVCVCTTSQS 60
DB 1 MSRLKAKPSLCKSEPLTTERVTTLSVLKRVITSCYGPSGRKLQHLNFGGVCVCTTSQS 60
QY 61 SALLSHLLVTHPILKILTASIQNHVSSFSDCGLFTAILCCNLIENTVQRLGTPPTVIRLN 120
DB 63 SALLSHLLVTHPILKILTASIQNHVSSFSDCGLFTAILCCNLIENTVQRLGTPPTVIRLN 120
QY 121 KHLLSLCISYLKSDTCGCRIPVDFSSQIILLCLVRSILTSKPACMLTRKETEHSALILR 180
DB 123 KHLLSLCISYLKSDTCGCRIPVDFSSQIILLCLVRSILTSKPACMLTRKETEHSALILR 180
QY 181 AFLTTPENAEAGHIIIGKSLIPLKQORVIDSTVLPGLIEMSEVOLMRLPIKSKTALK 240
DB 183 AFLTTPENAEAGHIIIGKSLIPLKQORVIDSTVLPGLIEMSEVOLMRLPIKSKTALK 240
QY 241 VALFCTTSLGSDTSGEGTVVSYGVSLNVLNLDQLNLGRQLISDHVDLVLCQKVIHPS 300

Db 241 VALFCTTSLGSDTSGTGVVSVGVLENVAVDQLLNLGRQLISDHVDLVLQKVIHPS 300
Qy 301 LKQFLNMHRIIADIRIGVTLMPELTKMTGTQPIGSLGSI CPNSVSGVKDVTAKFGSOHF 360
Db 301 LKQFLNMHRIIADIRIGVTLMPELTKMTGTQPIGSLGSI CPNSVSGVKDVTAKFGSKHF 360
Qy 361 FLHPNERTCSLLLCNRNDTAMDELKLTCTQALHVLQTLKBPWALLGGCTTETHLAAY 420
Db 361 FLHPNERTCSLLLCNRNDTAMDELKLTCTQALHVLQTLKBPWALLGGCTTETHLAAY 420
Qy 421 IRKTHNDPRLKDDCTQTELOLIAEAFCSALESVVGSLHDDGGHILTDKMGHLSV 480
Db 421 IRKTHNDPRLKDDCTQTELOLIAEAFCSALESVVGSLHDDGGHILTDKMGHLSV 480
Qy 481 QADSPCVANWPDLLSQCGGLYNSQBELNWSFLRSTRPFPVQSCFPHEAVSASNLTLTD 540
Db 481 QADSPCVANWPDLLSQCGGLYNSQBELNWSFLRSTRPFPVQSCFPHEAVSASNLTLTD 540
Qy 541 CLTAKLSGLQVAVETANLINDLSYVIEDKN 570
Db 541 CLTAKLSGLQVAVETANLINDLSYVIEDKN 570

RESULT 5

ADP43680
ID ADP43680 standard; protein; 498 AA.

XX AC ADP43680;

XX DT 18-NOV-2004 (first entry)

XX DE Human PMMM-37 protein SEQ ID NO:37.

XX KW human; protein modification and maintenance molecule; PMMM;
KW gastrointestinal; cardiovascular; immunosuppressive; antiinflammatory;
KW cytoskeletal; neuroprotective; gynaecological; gene therapy;
KW gastrointestinal disorder; cardiovascular disorder; autoimmune disorder;
KW inflammatory disorder; cell proliferative disorder;
KW developmental disorder; epithelial disorder; neurological disorder;
KW reproductive disorder.

XX OS Homo sapiens.

XX PN WO2004053068-A2.

XX PD 24-JUN-2004.

XX PF 03-DEC-2003; 2003WO-US038573.

XX PR 05-DEC-2002; 2002US-0431639P.

XX PR 17-DEC-2002; 2002US-0434315P.

XX PR 24-JAN-2003; 2003US-0442442P.

XX PR 30-JAN-2003; 2003US-0444141P.

XX PR 21-FEB-2003; 2003US-0449491P.

XX PA (INCY-) INCYTE CORP.

XX PI Becha SD, Hafalia AJA, Swarnakar A, Rannakar J, Richardson TW;

XX PI Kable AE, Marquis JP, Khare R, Lee SY, Tran UK, Bhatia UG;

XX PI Burrill JD, Blake JJ, Ho A, Zheng W, Gao J, Chawla NK, Mason PM;

XX PI Jin P, Lee S;

XX DR WPI; 2004-468843/44.

XX DR N-PSDB; ADP43721.

XX PT New human protein modification and maintenance molecules and
PT polynucleotides for diagnosing, preventing or treating diseases
PT associated with aberrant protein expression, e.g. cardiovascular or cell
PT proliferative disorders.

XX PS Claim 1; SEQ ID NO 37; 268pp; English.

XX CC The present sequence represents a human protein modification and

CC maintenance molecule (PMM) (I), designated PMMM-37. Also described: (1)
CC an isolated polynucleotide encoding (I); (2) a recombinant polynucleotide
CC comprising a promoter sequence operably linked to the polynucleotide in
CC (1); (3) a cell transformed with the recombinant polynucleotide; (4) a
CC transgenic organism comprising the recombinant polynucleotide; (5)
CC methods of producing or purifying (I); (6) an isolated antibody that
CC specifically binds to (I); (7) detecting a target polynucleotide or (I)
CC in a sample; (8) compositions comprising the polypeptide, an agonist
CC compound, an antagonist compound or an antibody, and an excipient; (9)
CC treating diseases or conditions associated with decreased expression or
CC overexpression of functional human PMM; (10) screening for a compound
CC that is effective as an agonist or antagonist of (I), that specifically
CC binds to (I), that modulates the activity of (I), or is effective in
CC altering expression of the target polynucleotide; (11) screening for a
CC potential toxicity of a test compound; (12) a diagnostic test for a
CC condition or disease associated with the expression of PMM in a
CC biological sample; (13) diagnosing a condition or disease associated with
CC the expression of PMM in a subject; (14) preparing a polyclonal or
CC monoclonal antibody with the specificity of the antibody in (6); (15) a
CC polyclonal or monoclonal antibody produced by the method in (14); (16)
CC compositions comprising the polyclonal or monoclonal antibody, and a
CC carrier; (17) generating an expression profile of a sample containing the
CC polynucleotides; and (18) an array comprising different nucleotide
CC molecules affixed at distinct physical locations on a solid substrate,
CC where at least one nucleotide molecule comprises a first oligonucleotide
CC or polynucleotide sequence specifically hybridizable with at least 30
CC contiguous nucleotides of the target polynucleotide. PMM sequences have
CC gastrointestinal, cardiovascular, immunosuppressive, antiinflammatory,
CC cytostatic, neuroprotective and gynaecological activities, and can be
CC used in gene therapy. The composition and methods are useful for
CC diagnosing, preventing or treating diseases or conditions associated with
CC aberrant expression of PMM, such as gastrointestinal, cardiovascular,
CC autoimmune/inflammatory, cell proliferative, developmental, epithelial,
CC neurological or reproductive disorders. They may also be used for
CC assessing the effects of exogenous compounds on the expression of nucleic
CC acid and amino acid sequences of PMM. The PMM or its fragments are also
CC useful in screening compounds for effectiveness as agonist or antagonist
CC of the polypeptides, or in altering the expression of the target
CC polynucleotide and compounds that specifically bind to or modulate the
CC activity of the polypeptide. The microarray is useful in monitoring or
CC measuring protein-protein interactions, drug-target interactions, and
CC gene expression profiles.

XX SQ Sequence 498 AA;

Query Match 84.6%; Score 2503; DB 8; Length 498;

Best Local Similarity 86.8%; Pred. No. 7.5e-251;

Matches 495; Conservative 2; Mismatches 1; Indels 72; Gaps 1;

Qy 1 MSRLKPKSLCKSEPLTTERVTTLSVLKRVTCYGPSGLKQLHNGFGGVCTTSQS 60

Db 1 MSRLKPKSLCKSEPLTTERVTTLSVLKRVTCYGPSGLKQLHNGFGGVCTTSQS 60

Qy 61 SALLSHLLVTHPILKILTASI QNHVSSFSDCLFTAILCCNLINENVRQLGTPPTVIRLN 120

Db 61 SALLSHLLVTHPILKILTASI QNHVSSFSDCLFTAILCCNLINENVRQLGTPPTVIRLN 120

Qy 121 KHLISLCISYLKSDTCGCRIPVDFTSTQILLCLIVRSILTSKPCMLTRKTEHVSAIILR 180

Db 121 KHLISLCISYLKSDTCGCRIPVDFTSTQILLCLIVRSILTSKPCMLTRKTEHVSAIILR 180

Qy 181 AFLLTIPENAEGHIIIGKSLIVPLKQVRVDTSTVPLGLIEMSEVOLMELLPIKKTALK 240

Db 181 AFLLTIPENAEGHIIIGKSLIVPLKQVRVDTSTVPLGLIEMSEVOLMELLPIKKTALK 240

Qy 241 VALFCTTSLGSDTSGTGVVSVGVLENVAVDQLLNLGRQLISDHVDLVLQKVIHPS 300

Db 241 VALFCTTSLGSDTSGTGVVSVGVLENVAVDQLLNLGRQLISDHVDLVLQKVIHPS 300

Qy 301 LKQFLNMHRIIADIRIGVTLMPELTKMTGTQPIGSLGSI CPNSVSGVKDVTAKFGSOHF 360

Db 301 LKQFLNMHRIIADIRIGVTLMPELTKMTGTQPIGSLGSI CPNSVSGVKDVTAKFGSKHF 360

QY 361 FHLLPNEATICSLLLCNRNDTAWDELKLTCTOTALHVLQTLTKEPWALLGGGCTETHLAAY 420
DB 361 FHLLPNEATICSLLLCNRNDTAWDELKLTCTOTALHVLQTLTKEPWALLGGGCTETHLAAY 420
QY 421 IIRKTHNDPESILKDDCTQTQLIAEAFCSALESVVGSLHDGGGELLTDMKYGHLWSV 480
DB 421 IIRK----- 424
QY 481 QADSPCVANWPDLLSCGCGLYNSQBELNWSFLRTRRPVPOSCLPHEAVGASNLITLD 540
DB 425 -----CGCGLYNSQBELNWSFLRTRRPVPOSCLPHEAVGASNLITLD 468
QY 541 CLTAKLSGLQVAVETANLIWDLVSIVIEDKN 570
DB 469 CLTAKLSGLQVAVETANLIWDLVSIVIEDKN 498

RESULT 6
ABG19636
ID ABG19636 standard; protein; 474 AA.
XX AC ABG19636;
XX 13-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #19627.
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US008631.
XX 31-MAR-2000; 2000US-00540217.
XX 23-AUG-2000; 2000US-00649167.
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
XX N-PSDB; AAS83823.
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX Claim 20; SEQ ID NO 4995; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this

CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
SQ Sequence 474 AA;
Query Match 78.6%; Score 2325; DB 4; Length 474;
Best Local Similarity 97.2%; Pred. No. 2.3e-232;
Matches 458; Conservative 4; Mismatches 7; Indels 2; Gaps 2;
QY 1 MSRLKAKPSLCKSEPLTTRVTRTILSVLKRIVTSCYGPSRLKQLHNGFGGVCTTSQS 60
DB 3 MSRLKAKPSLCKSEPLTTRVTRTILSVLKRIVTSCYGPSRLKQLHNGFGGVCTTSQS 62
QY 61 SALLSHLLVTHPILKILTASIQNHVSSPDCGLFTAILCCNLIENTVQRLGLTPTTVIRLN 120
DB 63 SALLSHLLVTHPILKILTASIQNHVSSPDCGLFTAILCCNLIENTVQRLGLTPTTVIRLN 122
QY 121 KHLISLCISYLSKSDTCGCRIPVDFSTOILLCLVRSILTSKPACMLTRKETHVSAIILR 180
DB 123 KHLISLCISYLSKSDTCGCRIPVDFSTOILLCLVRSILTSKPACMLTRKETHVSAIILR 182
QY 181 AFLTIPENAEGHIIILGKSLIVPLKQORVIDSTVLPGLIEMSEVOLMRLLPIKSTALK 240
DB 183 AFLTIPENAEGHIIILGKSLIVPLKQORVIDSTVLPGLIEMSEVOLMRLLPIKSTALK 242
QY 241 VALFCTTSLGSDTSGEGTVVSVGSLENVAVDQLNLNGLRQLISDHVDLVCQKVIHPS 300
DB 243 VALFCTTSLGSDTSXPGEGTVVSVGSLENVAVDQLNLNGLRQLISDHVDLVCQKVIHPS 302
QY 301 LKQFLNMHPIIADIRIGV-TLMEPL-TKMTGTQPIGSLGSI CPNSYGVSKDVCTAKFGSQ 358
DB 303 LKQFLNMHPIIADIRIGSDSDVEPLDLKWTGTQPIGSLGSI CPNSYGVSKDVCTAKFGSK 362
QY 359 HFFHLLIPNEATICSLLLCNRNDTAWDELKLTCTOTALHVLQTLTKEPWALLGGGCTETHLA 418
DB 363 HFFHLLIPNEATICSLLLCNRNDTAWDELKLTCTOTALHVLQTLTKEPWALLGGGCTETHLA 422
QY 419 AYIRKTHNDPESILKDDCTQTQLIAEAFCSALESVVGSLHDGGGELL 469
DB 423 AYIRKTHNDPESILKDDCTQTQLIAEAFCSALESVVGSLHDGGGELL 473

RESULT 7
AAB43205
ID AAB43205 standard; protein; 406 AA.
XX AC AAB43205;
XX 08-FEB-2001 (first entry)
XX Human ORFX ORF2969 polypeptide sequence SEQ ID NO:5938.
XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
XX vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
XX anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
XX immunosuppressant; thrombolytic; coagulant; vasotropic; antidiabetic;
XX hypotensive; dermatological; immunosuppressive; antiinflammatory;
XX antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
XX antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
XX neurodegenerative disorder; osteoarthritis; graft vs host disease;
XX cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
XX cholesterol ester storage; systemic lupus erythematosus; infection;
XX severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
XX allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
XX bone damage; cartilage damage; antiinflammatory disease; coagulation;
XX thrombosis; contraceptive.
XX Homo sapiens.
XX OS
XX WO200058473-A2.
XX PN
XX 05-OCT-2000.
XX PD

```
XX PF 31-MAR-2000; 2000WO-US008621.
XX PR 31-MAR-1999; 99US-0127607P.
XX PR 02-APR-1999; 99US-0127636P.
XX PR 03-APR-1999; 99US-0127728P.
XX PR 30-MAR-2000; 2000US-00540763.
XX PA (CURA-) CURAGEN CORP.
XX PI Shimkets RA, Leach M;
XX DE WPI; 2000-602362/57.
XX DR N-PSDB; AAC77414.
XX
XX Novel nucleic acids and peptides derived from open reading frame X,
XX useful for treating e.g. cancers, proliferative disorders,
XX neurodegenerative disorders and cardiovascular disease.
XX
XX Claim 11; Page 5123-5123; 5507pp; English.
XX
XX AAC74446 to AAC7606 encode the proteins given in AAB40237 to AAB43397,
XX which represent the human ORFX open reading frames 1 to 3161. The ORFX
XX sequences have activities such as: cytostatic; hepatotropic; vulnery;
XX antiproliferative; antiparkinsonian; nontropic; neuroprotective; osteopathic;
XX anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;
XX cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;
XX dermatological; immunosuppressive; antiinflammatory; antibacterial;
XX antiviral; antifungal; antirheumatic; antithyroid; and antianemic. The
XX sequences can be used for determining the presence of or predisposition
XX to, or preventing or treating pathological conditions associated with an
XX ORFX-associated disorder. The nucleic acids can be used to express ORFX
XX proteins in gene therapy vectors. The proteins and nucleic acids may be
XX used to treat cancers, proliferative disorders, neurodegenerative
XX disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
XX diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
XX storage, systemic lupus erythematosus, severe combined immunodeficiency
XX (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
XX disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
XX cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to
XX enhance coagulation; to inhibit thrombosis; and as a contraceptive
XX
XX Sequence 406 AA;
XX
XX Query Match 70.5%; Score 2086; DB 3; Length 406;
XX Best Local Similarity 99.0%; Pred. No. 1.4e-207;
XX Matches 402; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
XX
XX 165 MLTRKETEHSVALILRAFLITIPENAEHIIIGKSLIVPLKQGVVIDSTVLPGLIEMSE 224
XX |||||
XX 1 MLTRKETEHSVALILRAFLITIPENAEHIIIGKSLIVPLKQGVVIDSTVLPGLIEMSE 60
XX
XX 225 VQLMRLLPKKSTALKVAFCTTLTGSDTSGTGTVVSVGSLENAVLDQLNLGRQLI 284
XX |||||
XX 61 VQLMRLLPKKSTALKVAFCTTLTGSDTSGTGTVVSVGSLENAVLDQLNLGRQLI 120
XX
XX 285 SDHVDLVLCQVTHPSLKQFLNMRHIIADRIQVTLMEPLTKMTGTQPTIGSGISCPNSY 344
XX |||||
XX 121 SDHVDLVLCQVTHPSLKQFLNMRHIIADRIQVTLMEPLTKMTGTQPTIGSGISCPNSY 180
XX |||||
XX 345 GSVKDVCTAFKGSQHPFLHPIBNEATICSLLNCRNDTAWDELKLTCTQALHVLQTLKEP 404
XX |||||
XX 181 GSVKDVCTAFKGSQHPFLHPIBNEATICSLLNCRNDTAWDELKLTCTQALHVLQTLKEP 240
XX |||||
XX 405 WALLGGCGCTETHLAAYIRKHTNDPESILKDDCTOTELQLIAEAPCSALESVVGSLHND 464
XX |||||
XX 241 WALLGGCGCTETHLAAYIRKHTNDPESILKDDCTOTELQLIAEAPCSALESVVGSLHND 300
XX |||||
XX 465 GGEILDMKYGHLSVQADSPCVANWPDLLSQCGGLYNSQBELNWSFLRSTRPVPQOS 524
XX |||||
XX 301 GGEILDMKYGHLSVQADSPCVANWPDLLSQCGGLYNSQBELNWSFLRSTRPVPQOS 360
XX |||||
XX 525 CLPHEAVGSASNLTLDCLTAKLSGLQVAVETANLIWDLGYVIEDKN 570
XX |||||
XX 361 CLPHEAVGSASNLTLDCLTAKLSGLQVAVETANLIWDLGYVIEDKN 406
XX |||||
XX
XX RESULT 8
XX AAB56236
XX ID AAB56236 standard; protein; 222 AA.
XX AC AAB56236;
XX XX
XX DT 13-MAR-2001 (first entry)
XX XX
XX DE Human secreted protein sequence encoded by gene 142 SEQ ID NO:330.
XX XX
XX KW Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;
XX antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;
XX cerebroprotective; nontropic; neuroprotective; antibacterial; virucide;
XX fungicide; ophthalmological; gene therapy; pathological condition;
XX autoimmune disease; rheumatoid arthritis; hyperproliferative disorder;
XX neoplasm; cardiovascular disorder; cardiac arrest; cerebral ischaemia;
XX cerebrovascular disorder; angiogenesis; nervous system disorder;
XX Alzheimer's disease; infection; ocular disorder; corneal infection;
XX wound healing; skin aging; food additive; preservative; chromosome 20.
XX
XX OS Homo sapiens.
XX PN WO200070042-A1.
XX XX
XX PD 23-NOV-2000.
XX XX
XX PF 11-MAY-2000; 2000WO-US012788.
XX XX
XX PR 13-MAY-1999; 99US-0134068P.
XX XX
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX XX
XX PI Rosen CA, Ruben SM, Moore PA, Young PE, Komatsoulis GA, Birse CE,
XX Duan RD, Florence KA, Soppet DR;
XX XX
XX DR WPI; 2000-679828/66.
XX DR N-PSDB; AAC99977.
XX XX
XX PT Isolated nucleic acid molecule encoding a human secreted protein is used
XX in preventing, treating or ameliorating a medical condition.
XX
XX PS Claim 11; Page 1007; 1065pp; English.
XX XX
XX CC The polynucleotide sequences given in AAC99818 to AAC99977 encode the
XX human secreted proteins given in AAB56077 to AAB56362. Human secreted
XX proteins have activities based on the tissues and cells the genes are
XX expressed in. Examples of activities include: immunosuppressive;
XX antirheumatic; antirheumatic; antiproliferative; cytostatic; cardiant;
XX vasotropic; cerebroprotective; nontropic; neuroprotective; antibacterial;
XX virucide; fungicide; and ophthalmological. The human secreted
XX polynucleotides and proteins can be used to prevent, treat or ameliorate
XX a medical condition in e.g. humans, mice, rabbits, goats, horses, cats,
XX dogs, chickens or sheep. They are also used in diagnosing a pathological
XX condition or susceptibility to a pathological condition. Disorders which
XX are diagnosed or treated include autoimmune diseases e.g. rheumatoid
XX arthritis, hyperproliferative disorders e.g. neoplasms of the breast or
XX liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular
XX disorders e.g. cerebral ischaemia, angiogenesis, nervous system disorders
XX e.g. Alzheimer's disease, infections caused by bacteria, viruses and
XX fungi and ocular disorders e.g. corneal infection. The proteins can also
XX be used to aid wound healing and epithelial cell proliferation, to
XX prevent skin aging due to sunburn, to maintain organs before
XX transplantation, for supporting cell culture of primary tissues, to
XX regenerate tissues and in chemotaxis. The proteins can also be used as a
XX food additive or preservative to increase or decrease storage
XX capabilities. AAC99809 to AAC99817 and AAB56076 represent sequences used
XX in the exemplification of the present invention
XX
XX Sequence 222 AA;
```

Query Match 36.9%; Score 1092; DB 3; Length 222;
 Best Local Similarity 97.6%; Pred. No. 1.6e-104;
 Matches 206; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 360 FFLHPNEATICSLLLCNRNDTAWDELKLTCTQALHVLQTLKEPWALLGGGCTETHLAA 419
 DB 12 FFLFLMEATICSLLLCNRNDTAWDELKLTCTQALHVLQTLKEPWALLGGGCTETHLAA 71

QY 420 YIRKHTNDPESILKDDCTQTELQLIAEAFCSALESVVSGSLRHDGGEILTDKMYGHLMS 479
 DB 72 YIRKHTNDPESILKDDCTQTELQLIAEAFCSALESVVSGSLRHDGGEILTDKMYGHLMS 131

QY 480 VQADSPCVANWPDLSSQCGGLYNSQBELNWSFLRSTRFPVQSCPLPHEAVGSASNLTL 539
 DB 132 VQADSPCVANWPDLSSQCGGLYNSQBELNWSFLRSTRFPVQSCPLPHEAVGSASNLTL 191

QY 540 DCLTAKLSGLQVAVETANLILDSYVIEDKN 570
 DB 192 DCLTAKLSGLQVAVETANLILDSYVIEDKN 222

RESULT 9
 AAB56218
 ID AAB56218 standard; protein; 223 AA.
 XX
 AC AAB56218;
 XX
 13-MAR-2001 (first entry)
 XX
 XX Human secreted protein sequence encoded by gene 142 SEQ ID NO:312.
 XX
 KW Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;
 KW antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;
 KW cerebroprotective; neutropic; neuroprotective; antibacterial; virucide;
 KW fungicide; ophthalmological; gene therapy; pathological condition;
 KW autoimmune disease; rheumatoid arthritis; hyperproliferative disorder;
 KW neoplasm; cardiovascular disorder; cardiac arrest; cerebral ischaemia;
 KW cerebrovascular disorder; angiogenesis; nervous system disorder;
 KW Alzheimer's disease; infection; ocular disorder; corneal infection;
 KW wound healing; skin aging; food additive; preservative; chromosome 20.
 XX
 OS Homo sapiens.
 XX
 PN WO200070042-A1.
 XX
 PD 23-NOV-2000.
 XX
 PF 11-MAY-2000; 2000WO-US012788.
 XX
 PR 13-MAY-1999; 99US-0134068P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Ruben SM, Moore PA, Young PE, Komatsoulis GA, Birse CB;
 PI Duan RD, Florence KA, Soppet DR;
 XX
 XX WPI; 2000-679828/66.
 DR N-PSDB; AAC99959.
 XX
 PT Isolated nucleic acid molecule encoding a human secreted protein is used
 PT in preventing, treating or ameliorating a medical condition.
 XX
 PS Claim 11; Page 999-1000; 1065pp; English.
 XX
 CC The polynucleotide sequences given in AAC99818 to AAC99977 encode the
 CC human secreted proteins given in AAB56077 to AAB56362. Human secreted
 CC proteins have activities based on the tissues and cells the genes are
 CC expressed in. Examples of activities include: immunosuppressive;
 CC antiarthritic; antirheumatic; antiproliferative; cytostatic; cardiant;
 CC vasotropic; cerebroprotective; neutropic; neuroprotective; antibacterial;
 CC virucide; fungicide; and ophthalmological. The human secreted
 CC polynucleotides and proteins can be used to prevent, treat or ameliorate

CC a medical condition in e.g. humans, mice, rabbits, goats, horses, cats,
 CC dogs, chickens or sheep. They are also used in diagnosing a pathological
 CC condition or susceptibility to a pathological condition. Disorders which
 CC are diagnosed or treated include autoimmune diseases e.g. rheumatoid
 CC arthritis, hyperproliferative disorders e.g. neoplasms of the breast or
 CC liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular
 CC disorders e.g. cerebral ischaemia, angiogenesis, nervous system disorders
 CC e.g. Alzheimer's disease, infections caused by bacteria, viruses and
 CC fungi and ocular disorders e.g. corneal infection. The proteins can also
 CC be used to aid wound healing and epithelial cell proliferation, to
 CC prevent skin aging due to sunburn, to maintain organs before
 CC transplantation, for supporting cell culture of primary tissues, to
 CC regenerate tissues and in chemotaxis. The proteins can also be used as a
 CC food additive or preservative to increase or decrease storage
 CC capabilities. AAC99809 to AAC99817 and AAB56076 represent sequences used
 CC in the exemplification of the present invention
 XX
 SQ Sequence 223 AA;

Query Match 36.4%; Score 1076; DB 3; Length 223;
 Best Local Similarity 98.6%; Pred. No. 7.4e-103;
 Matches 213; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 165 MLTRKETEHSVALLIRAFLLTIPENAEGHIIILGKSLIVPLKQGVVDSTVLPGLIEMSE 224
 DB 1 MLTRKETEHSVALLIRAFLLTIPENAEGHIIILGKSLIVPLKQGVVDSTVLPGLIEMSE 60

QY 225 VQLMRLLPIKSKTALKVALFCTTLSGDTSDTGEGTVVSVGSLENVADQLNLGRQLI 284
 DB 61 VQLMRLLPIKSKTALKVALFCTTLSGDTSDTGEGTVVSVGSLENVADQLNLGRQLI 120

QY 285 SHQVDVLVLCQVTHPSLKQFLANHRIIADRGVTLMEPLTKMTGTQPTGSLGSI CPNSY 344
 DB 121 SHQVDVLVLCQVTHPSLKQFLANHRIIADRGVTLMEPLTKMTGTQPTGSLGSI CPNSY 180

QY 345 GSVKDVCTAKFGSQHFHFLIPNEATICSLLLCNRND 380
 DB 181 GSVKDVCTAKFGSKHFFHFLIPNXATICSLLLCNRNE 216

RESULT 10
 AAB56356
 ID AAB56356 standard; protein; 166 AA.
 XX
 AC AAB56356;
 XX
 13-MAR-2001 (first entry)
 XX
 XX Human secreted protein sequence encoded by gene 142 SEQ ID NO:450.
 XX
 KW Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;
 KW antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;
 KW cerebroprotective; neutropic; neuroprotective; antibacterial; virucide;
 KW fungicide; ophthalmological; gene therapy; pathological condition;
 KW autoimmune disease; rheumatoid arthritis; hyperproliferative disorder;
 KW neoplasm; cardiovascular disorder; cardiac arrest; cerebral ischaemia;
 KW cerebrovascular disorder; angiogenesis; nervous system disorder;
 KW Alzheimer's disease; infection; ocular disorder; corneal infection;
 KW wound healing; skin aging; food additive; preservative; chromosome 20.
 XX
 OS Homo sapiens.
 XX
 PN WO200070042-A1.
 XX
 PD 23-NOV-2000.
 XX
 PF 11-MAY-2000; 2000WO-US012788.
 XX
 PR 13-MAY-1999; 99US-0134068P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Ruben SM, Moore PA, Young PE, Komatsoulis GA, Birse CB;

PI Duan RD, Florence KA, Soppet DR;
XX WPI; 2000-679828/66.
XX
PT Isolated nucleic acid molecule encoding a human secreted protein is used
PT in preventing, treating or ameliorating a medical condition.
XX
PS Disclosure; Page 1060; 1065pp; English.
XX
XX The polynucleotide sequences given in AAC99818 to AAC99977 encode the
CC human secreted proteins given in AAB56077 to AAB56362. Human secreted
CC proteins have activities based on the tissues and cells the genes are
CC expressed in. Examples of activities include: immunosuppressive;
CC antithrombotic; antihemagic; antiproliferative; cytostatic; cardiant;
CC vasotropic; cerebroprotective; nootropic; neuroprotective; antibacterial;
CC virucide; fungicide; and ophthalmological. The human secreted
CC polynucleotides and proteins can be used to prevent, treat or ameliorate
CC a medical condition in e.g. humans, mice, rabbits, goats, horses, cats,
CC dogs, chickens or sheep. They are also used in diagnosing a pathological
CC condition or susceptibility to a pathological condition. Disorders which
CC are diagnosed or treated include autoimmune diseases e.g. rheumatoid
CC arthritis, hyperproliferative disorders e.g. neoplasms of the breast or
CC liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular
CC disorders e.g. cerebral ischaemia, angiogenesis, nervous system disorders
CC e.g. Alzheimer's disease, infections caused by bacteria, viruses and
CC fungi and ocular disorders e.g. corneal infection. The proteins can also
CC be used to aid wound healing and epithelial cell proliferation, to
CC prevent skin aging due to sunburn, to maintain organs before
CC transplantation, for supporting cell culture of primary tissues, to
CC regenerate tissues and in chemotaxis. The proteins can also be used as a
CC food additive or preservative to increase or decrease storage
CC capabilities. AAC99809 to AAC99817 and AAB56076 represent sequences used
CC in the exemplification of the present invention
XX
SQ Sequence 166 AA;

Query Match 28.4%; Score 841; DB 3; Length 166;
Best Local Similarity 99.4%; Pred. No. 1.4e-78;
Matches 163; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSRLKAKPSLCKSEPLTTRVTTLSVLKRIVTCYGPSGRLLKQLHNGFGGVCTTSQS 60
DB 3 MSRLKAKPSLCKSEPLTTRVTTLSVLKRIVTCYGPSGRLLKQLHNGFGGVCTTSQS 62

QY 61 SALLSHLLVTHPIKILKTASIQNHVSFSDGGLFTALCCNLNENVRIGLTPPTVIRLN 120
DB 63 SALLSHLLVTHPIKILKTASIQNHVSFSDGGLFTALCCNLNENVRIGLTPPTVIRLN 122

QY 121 KHLLSLCISYKSDTCGRIPVDFSSQTQILLCLVRSILTSKPAC 164
DB 123 KHLLSLCISYKSDTCGRIPVDFSSQTQILLCLVRSILTSKPAC 166

RESULT 11
AAG02893
ID AAG02893 standard; protein; 86 AA.
XX
AC AAG02893;
XX
XX 06-OCT-2000 (first entry)
DT
XX Human secreted protein, SEQ ID NO: 6974.
DE
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW Gene therapy; chromosome mapping.
XX
XX Homo sapiens.
XX
PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
XX 21-FEB-2000; 2000EP-00200610.

XX 26-FEB-1999; 99US-0122487P.
PR (GEST) GENSET.
XX
PA Dumas Milne Edwards J, Duciert A, Giordano J;
PI
XX WPI; 2000-500381/45.
XX N-PSDB; AAC02899.
DR
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX
PS Claim 13; SEQ ID NO 6974; 71pp + Sequence Listing; English.
XX
XX The present sequence is a polypeptide encoded by one of a large number of
CC 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs were
CC prepared from total human RNAs or polyA+ RNAs derived from 30 different
CC tissues. EST sequences usually correspond mainly to the 3' untranslated
CC region (UTR) of the mRNA because they are often obtained from oligo-dT
CC primed cDNA libraries. Such ESTs are not well suited for isolating cDNA
CC sequences derived from the 5' ends of mRNAs and even in those cases where
CC longer cDNA sequences have been obtained, the full 5' UTR is rarely
CC included. 5' ESTs are derived from mRNAs with intact 5' ends and can
CC therefore be used to obtain full length cDNAs and genomic DNAs. 5' ESTs
CC are also used in diagnostic, forensic, gene therapy and chromosome
CC mapping procedures. They are used to obtain upstream regulatory sequences
CC and to design expression and secretion vectors
XX
SQ Sequence 86 AA;

Query Match 14.6%; Score 432; DB 3; Length 86;
Best Local Similarity 98.8%; Pred. No. 1.8e-36;
Matches 85; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MSRLKAKPSLCKSEPLTTRVTTLSVLKRIVTCYGPSGRLLKQLHNGFGGVCTTSQS 60
DB 1 MSRLKAKPSLCKSEPLTTRVTTLSVLKRIVTCYGPSGRLLKQLHNGFGGVCTTSQS 60

QY 61 SALLSHLLVTHPIKILKTASIQNHVS 86
DB 61 SALLSHLLVTHPIKILKTASIQNHVS 86

RESULT 12
ADN18515
ID ADN18515 standard; protein; 538 AA.
XX
AC ADN18515;
XX
XX 02-DEC-2004 (first entry)
DT
XX
DE Bacterial polypeptide #1168.
XX
KW Recombinant DNA construct; transformed plant; improved plant property;
KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
KW pathogen tolerance; pest tolerance; plant disease resistance;
KW cell cycle pathway modification; plant growth regulator;
KW homologous recombination; seed oil yield; protein yield; carbohydrate;
KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
KW bacterial polypeptide.
XX
OS Bacteria.
XX
XX US2003233675-A1.
PN
XX
PD 18-DEC-2003.
XX
XX 20-FEB-2003; 2003US-00369493.
PF
XX 21-FEB-2002; 2002US-0360039P.
XX
XX

PA (CAOY/) CAO Y.
PA (HINK/) HINKLE G J.
PA (SLAT/) SLATER S C.
PA (CHEN/) CHEN X.
PA (GOLD/) GOLDMAN B S.
XX
XX Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
PI WPI; 2004-061375/06.
XX
XX New recombinant DNA construct comprising a promoter positioned to provide
PT for expression of a polynucleotide encoding a polypeptide from a
PT microbial source, useful for producing plants with improved properties.
XX
XX Claim 1; SEQ ID NO 1168; 122pp; English.
XX
XX The invention relates to a recombinant DNA construct comprising a
CC promoter functional in a plant cell, where the promoter is positioned to
CC provide for expression of a polynucleotide encoding a polypeptide from a
CC microbial source. The invention also relates to a transformed plant
CC comprising the recombinant DNA construct and a method of producing a
CC transformed plant having an improved property. The plant is a crop plant
CC such as maize or soybean. The method of producing a transformed plant
CC having an improved property comprises transforming a plant with the
CC recombinant DNA construct and growing the transformed plant, where the
CC polynucleotide or polypeptide is useful for improving plant properties.
CC The recombinant DNA construct is useful for producing plants with
CC improved plant properties, e.g. improved cold, heat or drought tolerance,
CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
CC increased resistance to plant disease, better growth rate by modification
CC of the cell cycle pathway with plant growth regulators, increased rate of
CC homologous recombination, modified seed oil or protein yield and/or
CC content, improved yield by modification of carbohydrate, nitrogen or
CC phosphorus use and/or uptake, by modification of photosynthesis or by
CC providing improved plant growth and development under at least one stress
CC condition, improved lignin production or improved galactomannan
CC production. This sequence represents a bacterial polypeptide used in the
CC scope of the invention. Note: The sequence data for this patent did not
CC form part of the printed specification but was obtained in electronic
CC format from USPTO at seqdata.uspto.gov/sequence.html.
XX
SQ Sequence 538 AA;

Query Match 8.7%; Score 257.5; DB 8; Length 538;
Best Local Similarity 24.1%; Pred. No. 4.8e-17;
Matches 115; Conservative 84; Mismatches 214; Indels 65; Gaps 19;
QY 17 LTTERTVTLVSLKRIVTCVGPGRLLKQLHNGFGVGVCTTSQSSALLSHLLVTHPIKI 76
DB 33 ILAETVRTL-----GPKGMDKMLVDSIGDIV-VTNDGVTLKEMDIEHPAAKM 80
QY 77 LTASIQNHVSSFDGCLFTAILCCNL-----ENVQRILGTPPTVIRLNKHLSLCISYLK 132
DB 81 LVEVAKTQDEVDG-GTTAVIAGELLKKAENLEIHP-TIIANGVRAQAEKAQEIL 138
QY 133 SDTCGCEIPVDFSTQILLCLVRSILTSKPACMLTRKTEHVSALILRAFLLTIPENAE- 191
DB 139 DD-----IAIDASDRDTLMKVAMTAMTKG-----TEKAREPLAELIVDA-VKQVEEDGEV 188
QY 192 --GHILGLKSLVPLKGORVIDSTVLPGLIEMSEVQLMRLLPIKSTALKVAFCTTLS 249
DB 189 EKDHIEKK-----EGAAVDDSTLVQGVIIKERVH--PGMP-KVVENKIALLCNPI- 239
QY 250 GDTSDTGEGTVVSYGVSLNAVLDQLNLGRQLISDHVD-----LVLCKQKVHPSLKQFL 305
DB 240 -EVKTEVDABEIRITDPSQMAFIEQEQEIMRDWNSIVDTGANVLCQKGIIDLAQHYL 298
QY 306 NMRIIADIRGVTLMELTKMTGTQPIGSLGICPNISYGVKDVCTAKP-GSQHFHLI 364
DB 299 AKAGVLAVRVKSDMEKLSKATGANIVTNIEDLSPEDLGEAGVWSEKKISGSEMFVEE 358
QY 365 PNEATICSLLLCNRNDTAWDELKLTCTALHVLQLTLKEPWWLLGGCGCTETHLAAVIRHK 424

DB 359 CKPEKAVTILVRGSTEHVHSEVERAIEDAIGVVAATVEDKVVAGGAPAEIEA----- 412
QY 425 TINDPESILKD--DECTOTELQLIABAFCSALESVGSLEHGG-----BILTMKYGH 476
DB 413 -----KRLKDYADSIISGRE-QLAVSAFAEALEIVPKTLAENAGLDSIDVLVDLRAH 463
RESULT 13
AAO24120
ID AAO24120 standard; protein; 560 AA.
XX
AC AAO24120;
XX
DT 11-MAR-2004 (first entry)
XX
XX S shibatae TF55 alpha subunit.
XX
KW Chaperonin; nanotemplate; nanostructure; nanosensor; nanodevice; TF55;
KW human; mouse; yeast.
XX
OS Sulfolobus shibatae.
XX
PN WO2003080796-A2.
XX
PD 02-OCT-2003.
XX
PF 08-NOV-2002; 2002WO-US035889.
XX
PR 08-NOV-2001; 2001US-0340538P.
XX
PA (USAS) NASA US NAT AERO & SPACE ADMIN.
PA (SETI-) SETI INST.
XX
PI Trent JD, Mcmillan RA, Kagawa H, Paavola CD;
XX WPI; 2003-845205/78.
XX
PT Novel chaperonin polypeptides, useful in field-effect transistor,
PT integrated circuit, field emitter, probe, nanocomputer, quantum computer,
PT sensor or biochip applications.
XX
PS Disclosure; Fig 15; 91pp; English.
XX
XX The present invention relates to an isolated polypeptide comprising an S.
CC shibatae TF55 beta subunit and its mutants. Also described are
CC nanotemplates and devices comprising at least one chaperonin protein such
CC as TF55. A nanodevice comprising at least one nanotemplate is used as a
CC field-effect transistor, integrated circuit, field emitter, probe,
CC nanocomputer, quantum computer or sensor. The nanodevice is a
CC nanoelectromechanical system. The nanodevice is an optoelectronic switch,
CC optical switch, lens, laser, photonic crystals, or amplifier. The
CC nanodevice is a magnetic memory device, a magnetic storage device, hard-
CC disk drive read heads, magnetic RAM, magnetic field sensor, magnetic
CC logic gates or a magnetic switch. The nanodevice is an electronic,
CC semiconductor, mechanical, electromechanical, magnetic, photonic,
CC optical, or optoelectronic device. The nanosensor is a multi-nanowell
CC assay plate, or a single-molecule probe for DNA detection and
CC hybridisation. The present sequence is a protein shown in the
CC exemplification of the invention
XX
SQ Sequence 560 AA;

Query Match 8.7%; Score 256.5; DB 7; Length 560;
Best Local Similarity 24.3%; Pred. No. 6.5e-17;
Matches 115; Conservative 84; Mismatches 221; Indels 53; Gaps 16;
QY 29 LKRIVTCVGPGRLLKQLHNGFGVGVCTTSQSSALLSHLLVTHPIKILTASTQNHVSSF 88
DB 31 LAEMURSSGPGKGLDKMLDSFGD-VTINDGATIVKMEIQHPAAKLIVKAAQDAEV 89
QY 89 SDGGLFTAILCCNLNIENVQRL---GLTPTTVIRLNKHLSLCISYLSKSTCGCRIPVDFS 145
DB 90 GDGTTSAVVLGALLEKAESLLDQNHPTIIIEGYKAVTKALELLPQ--LGTRIDRL 147

PA (CAOV/) CAO Y.
PA (HINK/) HINKLE G J.
PA (SLAT/) SLATER S C.
PA (CHEN/) CHEN X.
PA (GOLD/) GOLDMAN B S.
XX
XX
PI Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
XX
XX WPI; 2004-061375/06.
XX
XX New recombinant DNA construct comprising a promoter positioned to provide
PT for expression of a polynucleotide encoding a polypeptide from a
PT microbial source, useful for producing plants with improved properties.
XX
XX
XX Claim 1; SEQ ID NO 22860; 122pp; English.
XX
XX The invention relates to a recombinant DNA construct comprising a
CC promoter functional in a plant cell, where the promoter is positioned to
CC provide for expression of a polynucleotide encoding a polypeptide from a
CC microbial source. The invention also relates to a transformed plant
CC comprising the recombinant DNA construct and a method of producing a
CC transformed plant having an improved property. The plant is a crop plant
CC such as maize or soybean. The method of producing a transformed plant
CC having an improved property comprises transforming a plant with the
CC recombinant DNA construct and growing the transformed plant, where the
CC polynucleotide or polypeptide is useful for improving plant properties.
CC The recombinant DNA construct is useful for producing plants with
CC improved plant properties, e.g. improved cold, heat or drought tolerance,
CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
CC increased resistance to plant disease, better growth rate by modification
CC of the cell cycle pathway with plant growth regulators, increased rate of
CC homologous recombination, modified seed oil or protein yield and/or
CC content, improved yield by modification of carbohydrate, nitrogen or
CC phosphorus use and/or uptake, by modification of photosynthesis or by
CC providing improved plant growth and development under at least one stress
CC condition, improved lignin production or improved galactomannan
CC production. This sequence represents a bacterial polypeptide used in the
CC scope of the invention. Note: The sequence data for this patent did not
CC form part of the printed specification but was obtained in electronic
CC format from USPTO at seqdata.uspto.gov/sequence.html.
XX
SQ Sequence 557 AA;

Query Match 8.4%; Score 248.5; DB 8; Length 557;
Best Local Similarity 23.6%; Pred. No. 4.4e-16;
Matches 108; Conservative 85; Mismatches 214; Indels 51; Gaps 14;
QY 28 VLKRVTCYGPGRGLKQLHNGFGVYCTTSQSSALLSHLVTHPILKIITASIQNHVS 87
DB 36 VLAEMLKSSLGPRGLDMLVDAFGD-ITVNDGATIVKEMEIOHPAAKLLVEVAKAQDAE 94
QY 88 FSDCGLFTAILCNLIENVQL---GLTPTTVIRLNKHLISLCISYLSKSDTCGR----- 139
DB 95 VGDGTTVVVVLGALLEKAELDENLHPITIE-----GYTKAMEEARLVDEA 144
QY 140 -IPVDSSTOILCLVRSILTSKPACMLRKTEHVSALILRAFLITIPENAEQ--HIIIL 196
DB 145 AVPEVEDEDDSVLRRIAEFTLASK--FVGTGPERDKIISMVIDAIRTVAEKRPDGGYVDL 202
QY 197 GKSLIVPLKGQRTVDSTVLPGLIEMSEVQLMRLLPIKKSTALKVALFCTTISGDTSDTG 256
DB 203 DYVKIEKKGGSLDSKLVRGIVLDKEVH--PAMP-KRVENAKIIVLDAPLEVQRPDL- 258
QY 257 EGTVVVSYGVSLENVLDQ---LLNLGRQLISDHVDLVLCQKVHPSLKQFLNMHRIAI 313
DB 259 TTKIRVTDIEKLSFLEETEERMLRDMVEKIAATGANVVIQKGIDVQAQHFLLAKKGILAV 318
QY 314 DRIGVTLMELPTXMTGTQPIGSLGSCPNISYSGVKDVCYAKFGSQH--PFHLIPNEATIC 371
DB 319 RRVKRSIDIEKAVATGAKIVTSLRDLKPEYLGVAELVEERKVGEDKRVFTGAKNPKSV- 377
QY 372 SLLLCNRNDTAWDELKLTCTCTALHLQLTLKEPFWALLGGGCTETHLAAYIRHKHTNDPES 431

DB 378 TILLRGANDMLDDEARNIKDALHGLRNILREPKIVGGGAVEVELA----- 424
QY 432 ILKDECTQT---ELQLIABAFCSALESVVGSLHDGG 466
DB 425 -LKLKEFARTVGGKQQLAIEAYAEALETIPTVLAESAG 461
Search completed: September 3, 2005, 04:59:18
Job time : 173 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 6, 2005, 09:34:10 ; Search time 7919 Seconds
(without alignments)
11882.166 Million cell updates/sec

Title: US-10-616-263-29

Perfect score: 2472

Sequence: 1 tatgagccttcggaactgtg.....aaaaaaaaaaaaaaaaaa 2472

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: gb_est1:*

2: gb_est2:*

3: gb_hic:*

4: gb_est3:*

5: gb_est4:*

6: gb_est5:*

7: gb_est6:*

8: gb_gsa1:*

9: gb_gsa2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2361.2	95.5	2481	3	CR597193 full-leng
2	1544.2	62.5	2658	3	AK032528 Mus muscu
3	1544.2	62.5	2658	3	AK032554 Mus muscu
4	1198.2	48.5	2745	3	AK004987 Mus muscu
5	880.8	35.6	1105	4	BM473886 AGENCOURT
6	870.4	35.2	1019	4	BM558221 AGENCOURT
7	849.8	34.4	920	5	BQ641810 AGENCOURT
8	849.4	34.4	1137	1	AL532426 AL532426
9	843.4	34.1	930	5	BQ675098 AGENCOURT
10	797.8	32.3	932	5	BUS26879 AGENCOURT
11	785.4	31.8	810	4	BG742334 602631495
12	781	31.6	1044	5	BM913546 AGENCOURT
13	765	30.9	912	5	BQ423217 AGENCOURT
14	752.2	30.4	942	5	BX433357 BX433357
15	738	29.9	1013	1	AL532425 AL532425
16	729.4	29.5	737	4	BI085537 602870248
17	694.8	28.1	926	4	BG397725 602438786
18	693.8	28.1	697	5	BQ574234 UI-H-EZ1-
19	692.4	28.0	710	7	CN428047 170004243
20	674.2	27.3	879	7	CK467658 938975 WA
21	668.4	27.0	656	4	BG749677 602707221
22	663.4	26.8	929	5	BX452469 BX452469
23	660	26.7	722	2	BE279744 601157720
24	658.4	26.6	661	6	CA423280 UI-H-FE1-

25	656.2	26.5	764	4	BI549136
26	652.6	26.4	693	4	BI767302
27	648.8	26.2	652	5	BU686127 UI-CF-DUL
28	642.6	26.0	822	4	BM040705
29	642.4	26.0	644	5	BU947652 1055d10.x
30	641.8	26.0	803	7	CR753485 DKEPZp469H
31	638.2	25.8	887	4	BG036503 602326458
32	637.4	25.8	639	6	CB141647 K-EST0195
33	635.4	25.7	637	4	BM772823 K-EST0057
34	634.4	25.7	636	4	BM772193 K-EST0056
35	634.4	25.7	636	4	BM772822 K-EST0057
36	626.2	25.3	782	4	BI086889 602850311
37	623.8	25.2	627	5	BQ549598 ik88h03.x
38	623.4	25.2	760	2	BE386143 601277265
39	620.4	25.1	762	7	CR764167 DKEPZp468G
40	618.4	25.0	691	2	AW967111 EST379185
41	617.8	25.0	626	5	BM726839 UI-E-EJ0-
42	617.4	25.0	628	5	BU741135 UI-E-EJ0-
43	616.4	24.9	640	5	BX093992 BX093992
44	612.4	24.8	625	4	BG182054 RST917 AT
45	606.4	24.5	626	4	BG192494 RST11609

ALIGNMENTS

RESULT 1
CR597193
LOCUS full-length cDNA clone CS0DM014YF11 of Fetal liver of Homo sapiens (human).
DEFINITION
ACCESSION CR597193
VERSION CR597193.1 GI:50478000
KEYWORDS HTC; CDSIT_CDNA.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 2481)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
REMARK Contact : Feng Liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/ Invitrogen Corporation 1600 Faraday Avenue
REFERENCE 2 (bases 1 to 2481)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage : Bp 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen.
FEATURES
Location/Qualifiers
1..2481
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DM014YF11"
/tissue_type="Fetal liver"
/plasmid="pCMVSPORT_6"

ORIGIN
Query Match 95.5%; Score 2361.2; DB 3; Length 2481;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2363; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TATGAGCCTTCGGAACCTGTGGAGAGACTACAAAGTTTGGTTGTTATGTCCTTTAGT 60
|||||
DB 116 TATGAGCCTTCGGAACCTGTGGAGAGACTACAAAGTTTGGTTGTTATGTCCTTTAGT 175
|||||

QY 61 TGGGCTCATACATTTTGGGGTGGTACAGAAATCAAAAGCAGCCCTGTGTTTCCAAATACCTAA 120
Db 176 TGGGCTCATACATTTTGGGGTGGTACAGAAATCAAAAGCAGCCCTGTGTTTCCAAATACCTAA 235
QY 121 AAACGACGACATTCCTGAGCAAGATAGTCTGGGACTTTTCAAAATCTTTAGAGAGCCAAAT 180
Db 236 AAACGACGACATTCCTGAGCAAGATAGTCTGGGACTTTTCAAAATCTTTAGAGAGCCAAAT 295
QY 181 CCAGGGGAAGTAGCAGGCTTGCAATCTTCAGGTAAAGAGCAGCTTTTGAATCTGAGCTTC 240
Db 296 CCAGGGGAAGTAGCAGGCTTGCAATCTTCAGGTAAAGAGCAGCTTTTGAATCTGAGCTTC 355
QY 241 ATATCCAAAGAGAGATGAAATATACAGTTGAGATTAGAAAGAACCTGGCTTCTTGTAGCT 300
Db 356 ATATCGAAAGAGAGATGAAATATACAGTTGAGATTAGAAAGAACCTGGCTTCTTGTAGCT 415
QY 301 GGGATATCTTTTCATAGGTGTCCATCTTTGGAAATATCTTTTTCAGAGGTCTGCAAGAGCAG 360
Db 416 GGGATATCTTTTCATAGGTGTCCATCTTTGGAAATATCTTTTTCAGAGGTCTGCAAGAGCAG 475
QY 361 TCTGTAAATTTTCAGTCTCAAAGCAAAACAAAGAGATTTGAAGAGTGAAATGAAATTAAT 420
Db 476 TCTGTAAATTTTCAGTCTCAAAGCAAAACAAAGAGATTTGAAGAGTGAAATGAAATTAAT 535
QY 421 ATTTGGAAATTTACTAAATTTGTCAATTAATCATTTCTATGCTGATTAGCTTCAATAACATTGA 480
Db 536 ATTTGGAAATTTACTAAATTTGTCAATTAATCATTTCTATGCTGATTAGCTTCAATAACATTGA 595
QY 481 ACTTTTGTATTTTATAGCCCAATGCTGCATATTTCAATCTTTAAATCTTAAAGAAATTAAT 540
Db 596 ACTTTTGTATTTTATAGCCCAATGCTGCATATTTCAATCTTTAAATCTTAAAGAAATTAAT 655
QY 541 TTTAAATGTTTAAACGTTGATTAATGCAATAAATAGAAAAATGTGGTTTACAAAAATAAACG 600
Db 656 TTTAAATGTTTAAACGTTGATTAATGCAATAAATAGAAAAATGTGGTTTACAAAAATAAACG 715
QY 601 GTCTTCACATAGTTTACCACCTGAAGTAAGATGTCTCGTTTGGAGCTTGAAGGCCATCAT 560
Db 716 GTCTTCACATAGTTTACCACCTGAAGTAAGATGTCTCGTTTGGAGCTTGAAGGCCATCAT 775
QY 661 GTTAAAGAGTGAACCACTGACAACTGAGAGGTGAGGACCAACACTTTTCTGTCTTGAAAAAG 720
Db 776 GTTAAAGAGTGAACCACTGACAACTGAGAGGTGAGGACCAACACTTTTCTGTCTTGAAAAAG 835
QY 721 AATTGTAAACATCATGCTATGGGCCCTCAGGTAGGCTGAAAGCAGCTGCACAATGGCTTTGG 780
Db 836 AATTGTAAACATCATGCTATGGGCCCTCAGGTAGGCTGAAAGCAGCTGCACAATGGCTTTGG 895
QY 781 AGGTTACGTTGTACAACTCAGTCCCTCAGCTCTGCTCAGTCACTTTTGGTCAACA 840
Db 896 AGGTTACGTTGTACAACTCAGTCCCTCAGCTCTGCTCAGTCACTTTTGGTCAACA 955
QY 841 TCCATTTTAAAGATCCTGACAGCCCTCCATACAGAAATCATGTGCAAGCTTCAGTGATTG 900
Db 956 TCCATTTTAAAGATCCTGACAGCCCTCCATACAGAAATCATGTGCAAGCTTCAGTGATTG 1015
QY 901 TGGCTTATTCACAGCTATTTCTTGTGCAACCTTGATTGAAAAATGTTTACAGAGATTAGGCT 960
Db 1016 TGGCTTATTCACAGCTATTTCTTGTGCAACCTTGATTGAAAAATGTTTACAGAGATTAGGCT 1075
QY 961 GACACCCACACCTGCTATAGATTAAATAAACAATCTTTTGGCTTTTGGCATCAGTTATCT 1020
Db 1076 GACACCCACACCTGCTATAGATTAAATAAACAATCTTTTGGCTTTTGGCATCAGTTATCT 1135
QY 1021 CAACTCTGATACCTGTGGTGTGTCGAATCCAGGTGAGCTTTTAGTAGTACTCAGATCCTCCT 1080
Db 1136 CAACTCTGATACCTGTGGTGTGTCGAATCCAGGTGAGCTTTTAGTAGTACTCAGATCCTCCT 1195
QY 1081 TTGTTTGGTGGTGTATATTAACAAGTAAACCTGCTGTATGCTCACAGAAAGGAAAC 1140
Db 1196 TTGTTTGGTGGTGTATATTAACAAGTAAACCTGCTGTATGCTCACAGAAAGGAAAC 1255

QY 1141 AGAGCATGTCACTGCTTTTGATCCTGAGAGCCTTTTGTCTTACAATTCAGAAAAATGCTCGA 1200
Db 1256 AGAGCATGTCACTGCTTTTGATCCTGAGAGCCTTTTGTCTTACAATTCAGAAAAATGCTCGA 1315
QY 1201 AGGCCACATCATTTTATAGGAAGAGTTTAAATTTCTTACCTTTTAAAGGTCAAAGAGTTATAGA 1260
Db 1316 AGGCCACATCATTTTATAGGAAGAGTTTAAATTTCTTACCTTTTAAAGGTCAAAGAGTTATAGA 1375
QY 1261 TTCCACTGTATTAATCTCGGGATATCTCAATTTGAAATGTGAGAGTTCAATTAATGAGGCTATT 1320
Db 1376 TTCCACTGTATTAATCTCGGGATATCTCAATTTGAAATGTGAGAGTTCAATTAATGAGGCTATT 1435
QY 1321 ACCTATCAAAAAATCAACTGCCCTCAAGGTGCACTCTTTTGTACAACCTTTATCCGGAGA 1380
Db 1436 ACCTATCAAAAAATCAACTGCCCTCAAGGTGCACTCTTTTGTACAACCTTTATCCGGAGA 1495
QY 1381 CACTTCTGACACTGGAGAGAGAACTGTGTGTGTGCTGAGTTATGGGTTTTCTTTGAAAAATGC 1440
Db 1496 CACTTCTGACACTGGAGAGAGAACTGTGTGTGTGCTGAGTTATGGGTTTTCTTTGAAAAATGC 1555
QY 1441 AGCTTTGGACCAAGCTGCTTAACTAGGAAGGCAAGCTTAATCAGTGACCACTGTTGT 1500
Db 1556 AGCTTTGGACCAAGCTGCTTAACTAGGAAGGCAAGCTTAATCAGTGACCACTGTTGT 1615
QY 1501 CCTGTGCCAAAAGTTATACATCCATCTTTTGAAGCAGTTTCTCAATATGCTATGCTATTAT 1560
Db 1616 CCTGTGCCAAAAGTTATACATCCATCTTTTGAAGCAGTTTCTCAATATGCTATGCTATTAT 1675
QY 1561 TGCCATATAGACAGAAATTTGAGTGTGACTCTGATGGAACCCCTGACTAAAAATGACAGGAACACA 1620
Db 1676 TGCCATATAGACAGAAATTTGAGTGTGACTCTGATGGAACCCCTGACTAAAAATGACAGGAACACA 1735
QY 1621 GCCTATTTGGATCCCTAGGCTCAATATGCTTAATAGTTATGGAAGTGTGAAAGATGTGTG 1680
Db 1736 GCCTATTTGGATCCCTAGGCTCAATATGCTTAATAGTTATGGAAGTGTGAAAGATGTGTG 1795
QY 1681 CACTGCAAAATTTGGCTCCCAACATTTTTCATCTTATTCCTTAATGAAGCAACATCTG 1740
Db 1796 CACTGCAAAATTTGGCTCCCAACATTTTTCATCTTATTCCTTAATGAAGCAACATCTG 1855
QY 1741 CAGCTTGTCTTCTGCAACAGAAAAATGACACTGCTCCTGGGATGAGCTGAAGCTCACGTGTCA 1800
Db 1856 CAGCTTGTCTTCTGCAACAGAAAAATGACACTGCTCCTGGGATGAGCTGAAGCTCACGTGTCA 1915
QY 1801 GACGGCACTGCAATGTCTGCAAGTAACTCAAGGAAACCATGGGCTTTTGTGGGAGGTGG 1860
Db 1916 GACGGCACTGCAATGTCTGCAAGTAACTCAAGGAAACCATGGGCTTTTGTGGGAGGTGG 1975
QY 1861 CTGTACTGAAACTCAATTTGGCTGCAATATACAGACAGAACTCAACAGCACCAGAAAG 1920
Db 1976 CTGTACTGAAACTCAATTTGGCTGCAATATACAGACAGAACTCAACAGCACCAGAAAG 2035
QY 1921 CATTTCTCAAAGATGATGATGACTCAAAACAGAACTTCAATTAATTTGCTGAAGCAATTTTG 1980
Db 2036 CATTTCTCAAAGATGATGATGACTCAAAACAGAACTTCAATTAATTTGCTGAAGCAATTTTG 2095
QY 1981 CAGTGGCTTAGAATCTGTGTTGGCTCTTTTAGAAACATGATGAGAGTGAATTTCTCACTGA 2040
Db 2096 CAGTGGCTTAGAATCTGTGTTGGCTCTTTTAGAAACATGATGAGAGTGAATTTCTCACTGA 2155
QY 2041 CATGAAGTATGACACCTTTGGTGTGAGTTCCAGGCAAGATTCCTCCTGTGTTGCTAACTGGCC 2100
Db 2156 CATGAAGTATGACACCTTTGGTGTGAGTTCCAGGCAAGATTCCTCCTGTGTTGCTAACTGGCC 2215
QY 2101 AGATTGTCTTTCACAGTGTGGCTGTGGAATATACAAATAGCCAGGAAGAACTCAACTGGTTC 2160
Db 2216 AGATTGTCTTTCACAGTGTGGCTGTGGAATATACAAATAGCCAGGAAGAACTCAACTGGTTC 2275
QY 2161 TTTCTTAAAGAGACACGCTGTGCCATTTGTGTGCCAACAAAGCTGCCCTTCCACATGAAGCTGT 2220
Db 2276 TTTCTTAAAGAGACACGCTGTGCCATTTGTGTGCCAACAAAGCTGCCCTTCCACATGAAGCTGT 2335
QY 2221 GGGCTCAGCCAGCAACCTGTGACCTTGGACTGTTTGTGACTGCAAAAGCTTGTGGGCTACAGGT 2280


```

Db      2336 GGGCTCAGCAGCAGCACTGACCTGGACTGTGTTGACTGCAAGCTTAGTGCGCTTACAGGT 2395
Qy      2281 GGCTGTAGACAGACCCATTGATTTGGGATCTTTTCATATGTTATTGAAGATAAARACTA 2340
Db      2396 GGCTGTAGACAGACCAATTGATTTGGATCTTTTCATATGTTATTGAAGATAAARACTA 2455
Qy      2341 AGAAGATAGCATGTTCTGTATTACAG 2366
Db      2456 AGAAGATAGCATGTTCTGTATTACAG 2481

RESULT 2
LOCUS   AK032528
DEFINITION Mus musculus adult male olfactory brain cDNA, RIKEN full-length
           enriched library, clone:6430582D06 product:McKusick-Kaufman
           syndrome protein, full insert sequence.
ACCESSION AK032528.1 GI:26328342
KEYWORDS  HTC; CAP trapper.
SOURCE   Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
9279253
10349636
2
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20499374
11042159
3
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kiteunai, T., Tashiro, H., Itoh, M.,
Yamamoto, R., Matsumoto, S., Hazama, M., Nishine, T., Harada, A.,
Suzuki, R., Imoto, K., Ishii, Y., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsumura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-Format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
11076861
4
The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)
5
The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 2658)
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, M.,
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,

```

TITLE
JOURNAL

Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
Muramatsu, M. and Hayashizaki, Y.
Direct Submission
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.jp,
URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)
CDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.jp/
URL: http://fantom.gsc.riken.jp/
URL: http://Location/Qualifiers

FEATURES
source

1..2658
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:6430582D06"
/db_xref="taxon:10090"
/clone="6430582D06"
/sex="male"
/tissue_type="olfactory brain"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
759..2471
/note="unnamed protein product; McKusick-Kaufman syndrome
protein [MGJ|MGJ:1891836, GB|NM_021527, evidence: BLASTN,
99%, match=2657]
putative"
/codon_start=1
/protein_id="BAC27912.1"
/db_xref="GI:26328343"
/translation="MSRLBAKPSLCKTEPLTSEKVRSTLSVLKGVIASCYGSGRLK
QHNHGCGVYTTSSALLRNLSVTHPVKLITLSSVQNVHSCFSCGGLFTALCCNL
INERQLDITPATIKNLKYLISLCTSYLKSEACSCRIIPVDFRSTFLSLVHSILTS
KPCMLTRKETDHGAILKAFLLTIPSTETRMVIGKSIIVPLKQRTVDSVTLPL
LTAESVQLRLRLPTQKASGLVALFCTSLSGDFSNAGGVVVAHYQVLENVALEOL
NLHGRLVTHVDLVLCKVIHPSLKPESERHVMIDRVGVITLMESLSKVTCATPTG
SLNPVITYTSGVKDVCSPFSGKHFHLLPNEAVTCTLLCSNRDNTANEELKLTQCT
AMHVQLTITKEPWLIGGCTTHLAAVYVRHVEAEAIVRDDGTQAKLHVAAAEAF
CSALSESVAAGSLSHDGGELIDTKYHLNSCADSASVGNWSDTSLSCGGLVNSQBEL
SWSVLRSYTHPPAPQTCLPQAALGSASNLTVDCFYAKLSGLQVAVETANLILDSLYVI
EDKN"

CDS

polyA_signal 2639..2644
polyA_site 2658
/note="putative"
/note="putative"
ORIGIN

Query Match 62.5%; Score 1544.2; DB 3; Length 2658;
Best Local Similarity 80.3%; Pred. No. 2.4e+304;
Matches 1889; Conservative 0; Mismatches 448; Indels 16; Gaps 6;
Qy 1 TATGAGCCTTCGAACTTGTGGAGAGACTACAAAGTTTGGTGTATGCTCCCTTAGT 60
Db 147 TATGAGCTTTCAGATCTCTGGAGAGAGTACAAAGTTCTGATGTTATGATACATTAAT 206
Qy 61 TGGGCTCATACATTTGGGGTGGTGTACAGATCAAAAGCAGCCCTGTTTCCAAATACCTAA 120
Db 207 TGGGTTTCATACATTTGGGGTGGCAGAGATCAAAAGCCAGCCCTGCTTCTTCCAAGTCTCT-- 264
Qy 121 AAACGACGACATCTCTGACAGAGATAGTCTGGGACTTTCAAATCTTCAGAGAGCCCAAT 180
Db 265 -AAGGACGATACCACTGAGCTGGATAGTCTGGGACTTTCGAAGTCCACCGAGAGCAAC 323
Qy 181 CCAGGGGAAGTAGCAGGCTTGCAATCTTCAGGTAAAGAGCAGAGCTTGAATCTGAGCTTC 240

RESULT 3
AK032554
LOCUS
DEFINITION
AK032554 Mus musculus adult male olfactory brain cDNA, RIKEN full-length
enriched library, clone:6430593117 product:McKusick-Kaufman
syndrome protein, full insert sequence.
ACCESSION
VERSION AK032554.1 GI:26328364
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
99279253
10349636
2
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20499374
11042159
3
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kitesunai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Hashiwa, K.,
Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanishi, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
11076861
4
The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)
5
The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 2658)
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Saeki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
Muramatsu, M. and Hayashizaki, Y.
Direct Submission
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gsr.riken.jp,
URL: http://genome.gsc.riken.jp/; Tel: 81-45-503-9222,

Fax: 81-45-503-9216)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.jp/
URL: http://fantom.gsc.riken.jp/.

FEATURES
source
1. .2658
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:6430593117"
/db_xref="taxon:10090"
/clone="6430593117"
/sex="male"
/tissue_type="olfactory brain"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
759. .2471
/note="unnamed protein product; McKusick-Kaufman syndrome
protein (MGI:1891836, GB|NM_021527, evidence: BLASTN,
99%, match=2657)
putative"
/codon_start=1
/protein_id="BAC27921.1"
/db_xref="GI:26328365"
/translation="MSRLKAKPSLCKTEPLTSEKVRSTLSVLKGVASCYGSGRLK
QUNHGGVGVYTSOSSALLRNLSVTHPVKILTSVQNHVSCFSGGLFTALCNL
INIGRLDITPATIKLNKYLKLSCTSYLKSEACSCRI PVDPRSTHTFLSLVHSILTS
KPCMLTRKETDHIGALILKAFLLIPETERMVLGKSIILPLKQRTVDSVLFG
LLEASVQRRLFTQKASLRVALFCTSLSGDFSNAGGVVAVHVSLENAVLSQL
LNLGRLVTDHVDLVLCQKVIHPSKQFFSERHVAIDRVGTVMESLSKVGTATGPIG
SLNPVSTYTGKQVDCSARFGSKHFFHLHPNEATVCTLLGSRNDTAWELKLTQCT
AMHVLQATIKEPVLLGGCTTHLAAVVRHVKHVEAEI VRDDGCTQAKLHVAAAF
CSALRSVAGSLHEDGGEI LDITKYHLWSCQADSASVGNWSDTLSCGGLYNSBEL
SNVLSRSTYHPFPQTCFLPQAALGSASNLTVDFYAKLSGLQAVETANLILDSYVI
EDKN"
polyA_signal 2639. .2644
/note="putative"
polyA_site 2658
/note="putative"
ORIGIN
Query Match 62.5%; Score 1544.2; DB 3; Length 2658;
Best Local Similarity 80.3%; Pred. No. 2.4e-304;
Matches 1889; Conservative 0; Mismatches 448; Indels 16; Gaps 6;
Qy 1 TATGAGCCTTCGGAACCTTCTGGAGAGACTACAAAGTTTGGTGTATGGTCCCTTAGT 60
Db 147 TATGAGCCTTCGGAACCTTCTGGAGAGAGTACAAAGTTCTGATGGTTATGGTACCATTAAT 206
Qy 61 TGGGCTCATACATTTGGGGTGGTACAGAAATCAAAAGCAGCCCTGTTTCCAAATACCTAA 120
Db 207 TGGGTTTCATACATTTGGGGTGGCAGAGATCAAAACAGCCCTGTCTTCAAGTTCT-- 264
Qy 121 AAACGACGACATTCCTGAGCAAGATAGTCTGGGACTTTTCAAAATCTTCAGAAAGAGCCAAAT 180
Db 265 -AAGGACGATACCACTGAGCTGGATAGTCTGGGACTTTCAGAGTCCACACGAGAGCCAAAC 323
Qy 181 CAGGGGAGTAGTACAGGCTTGCATCTTCAGTAAAGAGCAGCTTTGATCTGAGCTTC 240
Db 324 CTAGGAGAGTAGTACCATCTTGCAGTTTTCAG--AAAGAGCAGCACTTTGATCTGAGCTTC 381
Qy 241 ATATCGAAGAGAGATGAAATACCACTTGGATTTAGAAAGCACTGGCTTCTTGTAGCT 300
Db 382 ATATTAAGAGAGAAATGAAACCCAGTTGGATTTAGAAAGCACTGGCTTCTTGTAGCT 441
Qy 301 GGGATATCTTTCATAGGTGTCCATCTTGGAAATACATCTTTTTCGAGAGGTCTCGCAAGACAG 360
Db 442 GGGGTGACTTTTCATAGGGTCCATCTTGGAAATACATCTTTATACAGAGAGTTCGAAAGAG 501

QY	361	TCTGTAATAATTTTCAGTCTCAAGCAAAACAAAGAGATATTGAAGAGTGAAGTAAATTAAT	420
Db	502	TCTGTGAGGTCTGAGGCCAGAGCAGACAAAGAAATATTGAAGATTAAGCAAAATAAAG	561
QY	421	ATTGGAATTAATAATTTGTCATTAATCAATCTATGCTGATTAGCTTCATAAACATTGA	480
Db	562	ACTTTGGATTATTA--TGTCATTAATTAATCTATGCTGTTATTTGTCGAAACCACTT	619
QY	481	ACTTTTGGATTTTATAGCCACAATGCTGCATATTCATCTTTAAATTCCTAAAGATAAT	540
Db	620	GATTTTGTAGTGGT-----AATGCTGCATGTGCAGAC--TCATACCTGGAGCCAGTC	671
QY	541	TTTAATGTTAAACGTGATAATGCAATAAATAGAAAATGTGGTTTACAAAATAAAAACG	600
Db	672	TTTAATGTTTCAACATGACAGTGCATAAATAAATGAAGCCATATGAACAACTAAAGCT	731
QY	601	GTCTTCACTAGTTTACCACCTGAAGTAAAGATGTCTCGTTTGGAGCTTAAGAGCCATCAT	660
Db	732	GCCTTCACAAAGTTA--CACCTGAAGCAGGATGTCTCGGTTAGAAAGCTTAAGAGCCATCGT	790
QY	661	GTGTAAAGATGAACCACTGACAACTGAGAGAGTCAGGACCACACTTTTCTGCTTGAAGA	720
Db	791	GTGTAAAACTGAACCAATGACAAAGTGAGAAAGTCAGGTCCACACTTTTCTGCTTGAAG	850
QY	721	AATTGTAACATCATGCTATGGGCCCTCAGGTAGGCTGAAGAGCTGCACAAATGGCTTTGG	780
Db	851	AGTCATAGCCCTCATGCTATGGCCCTTCAGGGAGGCTGAAGAGCTACACAATGGATTGG	910
QY	781	AGGTTACGTGTGTACAACTCAGCTCCTCAGCTCTGCTCAGTCAGTCACCTTTTGGTCACA	840
Db	911	AGGTTGTGTACACAACTCAGCTCCTCAGCCCTGCTTCGAAACCTTTTCAGTCACCCA	970
QY	841	TCCCATTTTAAAGATCCTGACAGCCCTCCATACAGAAATCATGTGTCAAGCTTCAGTGATT	900
Db	971	TCCCGTATTAAAGATCCTAAGGTCATCCGTGCAGAAATCAGTGTCTGCTTCAGTGACTG	1030
QY	901	TGGCTTATTACAGCTATTCTTTGCTGCAACCTGATGAAATGTTACAGAGATTAGGCTT	960
Db	1031	CGGCTTATTACAGCCATTTCTGTGCTGCAACTTTGATTGAATAATTTCAAGAGCTAGATT	1090
QY	961	GACACCCACACATGCTATTAGATTAATAATAACATCTTTGAGTCTTTGTCATCAGTTATCT	1020
Db	1091	GACACCCGCACTGCTATTAAATTAATAATAATACCTCTTGAGTCTCTGTACCAAGTTATCT	1150
QY	1021	CAAGTCTGATACCTGTGGTGTCCAAATCCAGTGGACTTTAGTGTGCTCAGATCCTCTCT	1080
Db	1151	CAAGTCTGAAGCCTGTAGTTGTCCGAATCCAGTTGACTTCAGAAAGTACACATACCTTCT	1210
QY	1081	TTGTTTGGTGGCTAGTATATAACAAGTAAACCTGCTGTATGCTCACCAAGAGAAAC	1140
Db	1211	CAGCTTGGTACACAGTATCTTAAACAAGCAACAGGCTGTATGCTCACCAAGAGAAAC	1270
QY	1141	AGAGCATGTGAGTCTTGTATCTGAGAGCCTTTTGTGCTTACAAATCCAGAAATGCTCGA	1200
Db	1271	AGATCATAGGTGCTTGTATTGTAAGCTTTTACTTACAAATCCAGAAAGTACAGA	1330
QY	1201	AGGCCACATCATTTTAGGAAAGAGTTTAAATGTACCTTTTAAAGAGTCAAGAGATTATAGA	1260
Db	1331	AGAACGAATGGTTTTAGGGAAGAGTATATTGTTCCTTTAAAGGGCCAAAGAGTTACAGA	1390
QY	1261	TTCCACTGTATTACCTGGGATCTCATTTGAAATGTGAGAAATGTCAGAGTTCAATTAATGAGGCTATT	1320
Db	1391	TTCTACTGTATTACCTGGACTACTCATTTGAAGCATCAGAAATTCAAATTAAGGAGGTTATT	1450
QY	1321	ACCTATCAAAAAATCAACTGCCCTCAAGGTGGGCACTCTTTTGTACAACTTTATCCGGAGA	1380
Db	1451	ACCCACTCAAAAGGCAAGTGGCCTTAGGGTGGCACTCTCTGTACATCTTTATCTGGAGA	1510
QY	1381	CACCTCTGACACTGGAGAGGAACCTGTGGTGGTCAGTTATGGGGTTTCTCTTGAATAATGC	1440
Db	1511	TTTCTTCTAATGCTGGAGAGGAGTGGTGGTGGGCCCATTTACCAAGTGTCCCTTGGAGATGC	1570
QY	1441	AGTCTTGGACCACTGCTTTAACTAGGAGGAGCAGCTAATCAGTAGCCAGCAGTAGATCTTCT	1500
Db	1571	AGTTTTAGAGCAGTTGCTTTAATCTAGGAAGGCGACTAGTCACTGATCATGTAGACCTTGT	1630
QY	1501	CCTGTGCCAAAAAGTTATACATCCATCTTTTGAAGCAGTTTCTCAATATGCTATGTTAT	1560
Db	1631	CCTGTGCCAAAAAGTTAAATCCACCCATCTTTGAAACAGTTCTTTCAGTGAGCGCACGTTAT	1690
QY	1561	TGCCATAGACAGAAATGGAGTGACTCTGTGATGAAACCCCTGACTAAATGACAGGAACACA	1620
Db	1691	GGCCATCGACAGATTTGGGGTGACTCTGTATGAAATCTCTAAGCAAAAGTGACAGGACAA	1750
QY	1621	GCCTATTGGATCCCTAGGCTCAATATGCTCAATAGTTATGAAAGTGAAGAGATGTG	1680
Db	1751	GCCTATTGGTTCTCTAAACCCCAATAGTTTCTACTACTTATGGAAGTGTGAAGATGTG	1810
QY	1681	CACTGCAAAATTTGGCTCCCAACTTTTTTTCATCTTATTCCTAATGAAGCAACAATCTG	1740
Db	1811	CTCTGCAGATTTGGCTCCAAACATTTTTTTCATCTTCTTCTTAATGAGCCACTGTCTG	1870
QY	1741	CAGCTTGTCTTCTGCAACAGAAATGACACTGCTGGGATGAGCTGAAGCTCACGTGTCA	1800
Db	1871	CACCTTGTCTCTGCGAGCAGAAATGACACCGCTGGGAGGAGCTGAAGCTCACATGTCA	1930
QY	1801	GAGGCACTGCATGCTCTGCACTTAACACTCAAGNAACCATGGGCTTTGTTGGGAGGTGG	1860
Db	1931	AAACAGCAATGTCAGCTTTCAGTTTAACTAAGGAACCGTGGGTTTATTGGGAGGTGG	1990
QY	1861	CTGTACTGAACTCATTTTGGCTGTCATATATCAGACACAAGACTCAACAGCCACCAAGAA	1920
Db	1991	CTGTACAGBAACACACTTGGCTGCATATCTCAGACACAAGGTTTCATCAGAGGCAGAGC	2050
QY	1921	CATTCTCAAAGATGATGAATGACTCAACAGAACTTCAATTAATGCTGGAAGCAATTTG	1980
Db	2051	TATTGTGAGAGATGATGGGTGACTCAGGCAAAAGCTGCATGTGCTGTGAAGCAATTTG	2110
QY	1981	CAGTCCCTTAGAATCTGTTGTTGGCTCTTTAGAAATGATGAGGAGTGAATTTCTCACTCA	2040
Db	2111	CAGTGTCTGGAGTCCGTTGCTGGCTCTTTTGGAAATGATGTTGGTGAATCCTCATTTGA	2170
QY	2041	CATGAAGTATGACACACTTTTGGTCAAGTTTCAGGCAGATTTCTCCCTGTGTTGCTAACTGGCC	2100
Db	2171	CACGAAGTATGACACACTTTGCTCTGCAAGCAGATTTCTGCTCTGTTGGTAACTGGTC	2230
QY	2101	AGATTTGCTTTCAGAGTGTGGCTGTGGAATTAACAATAGCAGGAAGAACTCAACTGGTC	2160
Db	2231	AGATACGCTGTCCAGGTGTGGCTGTGGTTTGTACCAACAGCCAGGAAGAGCTCAGCTGGTC	2290
QY	2161	TTTCTTAAAGAAGCACAGCTCGTCCATTTGTGTCACAAAGCTGCCCTTCCACATGAACTGT	2220
Db	2291	TGTCTTAAAGATGACTTATCATCTTTTGCACACAACCTGCCCTTCCACAGGAGCTTT	2350
QY	2221	GGGCTCAGCCAGCAACCTTGACCTTGGACTGTTTGAAGCTGCAAAAGCTTTAGTGGCTACAGGT	2280
Db	2351	GGGCTCAGCCAGTAACTGACTGTGGACTGCTTCACTGCAAGCTGAGTGGCTTACAGT	2410
QY	2281	GGCTGTAGAGACAGCCAAATTTGAATTTGGGATCTTTCATATGTTATTGAAGATAAAACTA	2340
Db	2411	GGCTGTAGAGACAGCCAAATTTGAATTTTAGATCTTTTATATGTCTTTCATATGTGAAGATAAAACTA	2470
QY	2341	AGAGATAGCATG 2353	
Db	2471	ACATAAGAGAAATG 2483	
RESULT 4			
AK004987			
LOCUS			
DEFINITION			
AK004987			
ACCESSION			
VERSION			
AK004987.1			
GI:12836588			
2745 bp			
mRNA			
linear			
HTC 03-APR-2004			
Mus musculus adult male liver cDNA, RIKEN full-length enriched			
library, clone:1300013B18 product:McKusick-Kaufman syndrome			
protein, full insert sequence.			
AK004987			

KEYWORDS SOURCE ORGANISM

HTC; CAP trapper.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1 Carninci, P. and Hayashizaki, Y.

High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)

99279253

10349636

REFERENCE

2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)

20499374

11042159

REFERENCE

3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)

20530913

11076861

REFERENCE

4 The RIKEN Genome Exploration Research Group Phase II Team and the
PANTOM Consortium
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)

REFERENCE

5 The PANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)

6 (bases 1 to 2745)

REFERENCE

Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Arai, A., Aono, H.,
Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Fukunishi, Y.,
Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K.,
Hiraoka, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Izawa, M.,
Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K.,
Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C.,
Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D.,
Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y.,
Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T.,
Tejima, Y., Toyota, T., Yamamura, T., Yasunishi, A., Yoshida, K.,
Yoshino, M., Muramatsu, M. and Hayashizaki, Y.
Direct Submission

TITLE

Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Teurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.jp,
URL: http://genome.gsc.riken.jp/; Tel: 81-45-503-9222,
Fax: 81-45-503-9216)

COMMENT

Please visit our web site (<http://genome.gsc.riken.jp/>) for further
details.
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues. First strand cDNA was primed with a primer
15' GAGAGAGAGCGCGCACTCGAGTTTCTTTTCTTTTNN 3', cDNA was
prepared by using trehalose thermo-activated reverse transcriptase

and subsequently enriched for full-length by cap-trapper. Second
strand cDNA was prepared with the primer adapter of sequence 5'
GAGAGAGAGAGTCCCAAGCTCAATTAATTAATAAACCCCCCC 3'. cDNA was
cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end:
XhoI. Host: SOUR.

FEATURES

source

Location/Qualifiers
1..2745
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="PANTOM DB:1300013E18"
/db_xref="taxon:10090"
/clone="1300013E18"
/sex="male"
/tissue type="liver"
/clone lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
754..2262

CDS

/note="unnamed protein product; McKusick-Kaufman syndrome
protein (MGD|MG1:1891836, GB|NM_021527, evidence: BLASTN,
99%, match=2657)
putative"
/codon_start=1
/protein_id="BAB23722.1"
/db_xref="GI:12836589"
/translation="MSRLKAKPLSKTEPLTSEKVRSTLSVLKGVASCYGSPGRKL
QLHGLGCVYTTSSALLRSLVTHPVKILTSVQNHVSCFSDGLFTALCCNL
ISNIQLDLTPATYKLNKYLSSICTSYLKSACSCRI PVDPRSTHTPTLSLHSLITS
KPCAMLTRKETHIGALILKAFLLIPESTERMVLGKSIIVPLGKQRTDSTVLFG
LIEASEVQLRLLPTQKPFSEFNRVMAIDRVGVLMESLKSQVDTGSLNSPLIVSTY
GSVDVCSARFGSKHFPILLPNEATCTLLCSRNDTAWEEKLKTCTQAMHVLQLTIK
EPWLLGGCTETHLAAVYRHKVHEAEAVRDDGCTQAKLHVAEAFCSALESVAGS
LHSDGEILIDTKYGLHWSQADSASVGNWSDTLRSCGGLYNSQSELSVLRSTVH
PPAPQCLPQALGSANLTVDCFTAKLSGLQVAVETANLILDLISVIEDKN"
2724..2729
polya_signal
/note="putative"
2745
polya_site
/note="putative"

ORIGIN

Query Match 48.5%; Score 1198.2; DB 3; Length 2745;
Best Local Similarity 73.5%; Pred. No. 9.1e-234;
Matches 1730; Conservative 0; Mismatches 403; Indels 220; Gaps 7;
QY 1 TATGAGCCTTCGGAACCTTCTGGAGAGACTACAAAGTTTGTGTTGTTGTCCTCTTGTAGT 60
DB 142 TATGAGCTTTCAGATCTCTGGAGAGAGTACAAAGTTCTGATGTTTATGTCACCAATAT 201
QY 61 TGGGCTCATACATTTGGGCTGGTACAGAAATCAAAAGCAGCCCTGTGTTTCCAAATACCTAA 120
DB 202 TGGGTTTCATACATTTGGGCTGGCAGAGATCAAAAGCAGCCCTGTCTTCCAAGTTCT-- 259
QY 121 AAACGACGACATCTCTGAGCAAGATAGTCTGGGACTTTCAAATCTTCAGAAGAGCCAAAT 180
DB 260 -AAGGACGATACACTGAGCTGGATAGTCTGGGACTTTCGAAGTCCACCAAGAGCCAAAC 318
QY 181 CCAGGGGAGTAGCAGGCTTGCAATCTTCAGGTAAAGAGCAGCTTTGATCTCGAGCTTC 240
DB 319 CTAGAGAGAGTACCATCTTCAGCTTTTTCAG--AAAGAAGCAACTTTGATCTCGAACTTG 376
QY 241 ATATCGAAGAAGAGATGAAAATACCAGTTGGATTTAGAAAAGAACTGGCTTCTTTGTAGCT 300
DB 377 ATAATAAAGAAGAAATGAAAACCCCACTTGGATTGAGAAAGAACTGGCTTCTTTGTGGCT 436
QY 301 GGGATATCTTTTCATAGGTGTCATCTTGGAACATACCTTTTTCAGAGAGTCTGCAAGAGC 360
DB 437 GGGGTGACTTTTCATAGGCGTCCATCTTGGAACTATCTTTATACAGAGAGTTGCAAAAGAG 496
QY 361 TCTGTAAAAATTTTCAGTCTCAAGCAACAAAGAGATTTTGAAGAGTGAAGTGAAGTAAAAAT 420
DB 497 TCTGTAGGGCTTGAGCCAGAGGAGCAACAAAGAAATATTGAAGAATGAAGCAAAATAAG 556
QY 421 ATTTGGAATTACTAATTTTGTCTAATTAATCTATGCTGATTAGTCTTCAATAAATGGA 480

Query Match		35.2%;	Score 870.4;	DB 4;	Length 1019;
Best Local Similarity		97.0%;	Pred. No. 6.2e-167;		
Matches 908;		Conservative 0;	Mismatches 26;	Indels 2;	Gaps 2;
QY	844	CATTTTAAAGATCCTGACAGCCTCCATACAGAAATCATGTGTCAAGCTTCAGTGATTGTGG	903		
DB	1	CATTTTAAAGATCTGACAGCCTCCATACAGAAATCATGTGTCAAGCTTCAGTGATTGTGG	60		
QY	904	CTTATTACAGCTATTCTTTGCTGCAACCTGATGAAATGTTACAGATTAGGCTTGAC	963		
DB	61	CTTATTACAGCTATTCTTTGCTGCAACCTGATGAAATGTTACAGATTAGGCTTGAC	120		
QY	964	ACCACACATGCTATTAGATAAATAAATCAATCTTTGAGTCTTTGATCATGATTCTCAA	1023		
DB	121	ACCACACATGCTATTAGATAAATAAATCAATCTTTGAGTCTTTGATCATGATTCTCAA	180		
QY	1024	GTCTGATACCTGTGGTGTGCGAATCCAGTGGACTTTAGTAGTACTCAGATCCTCTTTG	1083		
DB	181	GTCTGAGACTGTGGTGTGCGAATCCAGTGGACTTTAGTAGTACTCAGATCCTCTTTG	240		
QY	1084	TTTGGTGGGAGTATATTAAACAGTAAACCTGCTGTATGCTCACCGAAGGAAACAGA	1143		
DB	241	TTTGGTGGGAGTATATTAAACAGTAAACCTGCTGTATGCTCACCGAAGGAAACAGA	300		
QY	1144	GCATGTCAGTCTTGTGATCTGAGAGCCTTTTGTCTTCAATTCAGAAATGCTGAAGG	1203		
DB	301	GCATGTCAGTCTTGTGATCTGAGAGCCTTTTGTCTTCAATTCAGAAATGCTGAAGG	360		
QY	1204	CCACATCATTTTAGGAAAGATTTAATTTGATCTTTTAAAGGTCAAGAGTTATAGATTTC	1263		
DB	361	CCACATCATTTTAGGAAAGATTTAATTTGATCTTTTAAAGGTCAAGAGTTATAGATTTC	420		
QY	1264	CACGTATTACCTGGGATCTCATTTGAAATGTCAGAAAGTTCAATTAATGAGGCTATTACC	1323		
DB	421	CACGTATTACCTGGGATCTCATTTGAAATGTCAGAAAGTTCAATTAATGAGGCTATTACC	480		
QY	1324	TATCAAAAATCACTGCCCTCAAGTGGCACTTTTGTGACAACTTTATCCGGAGACAC	1383		
DB	481	TATCAAAAATCACTGCCCTCAAGTGGCACTTTTGTGACAACTTTATCCGGAGACAC	540		
QY	1384	TTCTGACATCTGGAGAGGAATCTGCTGTGTCAGTATGGGTTTCTCTTGAAATGCAGT	1443		
DB	541	TTCTGACATCTGGAGAGGAATCTGCTGTGTCAGTATGGGTTTCTCTTGAAATGCAGT	600		
QY	1444	CTTGGACCAAGCTCTTAACTAGGAAGGCAAGCTTAATCAGTGACCAAGTATCTTCTCT	1503		
DB	601	CTTGGACCAAGCTCTTAACTAGGAAGGCAAGCTTAATCAGTGACCAAGTATCTTCTCT	660		
QY	1504	GTGCCAAAAGTTATACATCCATCTTTGAGCAGTTTCTCAATATCATCGTATTATTGC	1563		
DB	661	GTGCCAAAAGTTATACATCCATCTTTGAGCAGTTTCTCAATATCATCGTATTATTGC	720		
QY	1564	CATAGACAGAAATGGGAGTGAATCTGATGGAACCCCTGACTTAAATGACAGGACACAGC	1623		
DB	721	CATAGACAGAAATGGGAGTGAATCTGATGGAACCCCTGACTTAAATGACAGGACACAGC	780		
QY	1624	TATTGGATCCCTAGGCTCAATATGCTTAATAGTTATGGAAGTGTCAAGATGTGTCAC	1683		
DB	781	TATTGGATCCCTAGGCTCAATATGCTTAATAGTTATGGAAGTGTCAAGATGTGTCAC	840		
QY	1684	TGCAAAATTTGGCTCCCAACATTTTTTTCAT-CTTATTCTTAAT-GAAGCAACATCTGC	1741		
DB	841	TGCAAAATTTGGCTCCCAACATTTTTTTCAT-CTTATTCTTAAT-GAAGCAACATCTGC	900		
QY	1742	AGCTTGCTCTCTGCAACAGAAATGACACTGCTGCTGG 1777			
DB	901	AAGCTTGCTCTCTCTGCAACAAAAATGACCCCGG 936			

RESULT 7
BQ641810
LOCUS

DEFINITION	AGENCOURT_8290193 NIH_MGC_43 Homo sapiens cDNA clone IMAGE:6305558		
ACCESSION	5', mRNA sequence.		
VERSION	BQ641810		
KEYWORDS	BQ641810.1 GI:21765982		
SOURCE	EST.		
ORGANISM	Homo sapiens (human)		
REFERENCE	Homo sapiens		
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
TITLE	Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.		
JOURNAL	NIH-MGC http://mgc.nci.nih.gov/.		
COMMENT	1 (bases 1 to 920) National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: c9apbs-r@mail.nih.gov Tissue Procurement: ATCC cDNA Library Preparation: Rubin Laboratory DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL) Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLCM2528 row: a column: 15 High quality sequence stop: 625.		
FEATURES	Location/Qualifiers		
SOURCE	1..920 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="IMAGE:6305558" /tissue_type="normal pigmented retinal epithelium" /lab_host="DH10B (phage-resistant)" /clone_lib="NIH_MGC_43" /note="Organ: eye; Vector: pOTB7; Site 1: XhoI; Site 2: ECORI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library. "		
ORIGIN	Query Match 34.4%; Score 849.8; DB 5; Length 920; Best Local Similarity 97.2%; Pred. No. 9.8e-163; Matches 896; Conservative 0; Mismatches 23; Indels 3; Gaps 3;		
QY	1249	AAGAGTTATAGATTCCACTGTATTACCTGGGATCTCAATTGAAATGTCAGAAAGTTCAAT	1308
DB	1	AAGAGTTATAGATTCCACTGTATTACCTGGGATCTCAATTGAAATGTCAGAAAGTTCAAT	60
QY	1309	AATAGGCTTATACCTATCAAAAAATCAACTGCCCTCAAGTGGCACTCTTTTGTACAAC	1368
DB	61	AATAGGCTTATACCTATCAAAAAATCAACTGCCCTCAAGTGGCACTCTTTTGTACAAC	120
QY	1369	TTTATCCGAGACACTTCTGACACTGGGAGAGCACTGTGTGTCAGTTATGGGTTTC	1428
DB	121	TTTATCCGAGACACTTCTGACACTGGGAGAGCACTGTGTGTCAGTTATGGGTTTC	180
QY	1429	TCCTGAAAATGACAGTCTTGGACCACTGCTTAACTAGGAAGCAGCTAATCAGTGACCA	1488
DB	181	TCCTGAAAATGACAGTCTTGGACCACTGCTTAACTAGGAAGCAGCTAATCAGTGACCA	240
QY	1489	CGTAGATCTTGTCTGTGCAAAAAAGTTATACATCCATCTTTGAAGCAGTTTCTCAATAT	1548
DB	241	CGTAGATCTTGTCTGTGCAAAAAAGTTATACATCCATCTTTGAAGCAGTTTCTCAATAT	300
QY	1549	GCATGCTATTATTTGCCATAGACAGAAATGGAGTGACTCTGTATGGAAACCCCTGACTAAAAT	1608
DB	301	GCATGCTATTATTTGCCATAGACAGAAATGGAGTGACTCTGTATGGAAACCCCTGACTAAAAT	360
QY	1609	GACAGGAACACAGCCTATTGGATCCCTAGGCTCAATATATGTCCTTAATAGTTATGGAAAGTGT	1668

Db	957	CAWCCCATTTTAAKATCTWKAYAK-CWACAWACAACAATCATKTKTAKCAAKCWTCAKTKAT	1015
Qy	899	TGTGGCTTATTACAGCTATTCTTTGCTGCTCAACCTGATTGAAAATGTTTCAGAGATTAGGC	958
Db	1016	TKTKTCTAAT--CAGAAGCATCTTTGCTKCAACATGATK--AAATKTCAGAGATATGYT	1071
Qy	959	TTGACACCCACCACTGTCATTTAGATTAAATA	989
Db	1072	TKCAMCMCACTCATATGWTAAAAAAAHHM	1102
RESULT 9			
LOCUS	BQ675098	930 bp mRNA linear	EST 15-JUL-2002
DEFINITION	AGENCOURT 8349137 NIH_MGC_102 Homo sapiens cDNA clone IMAGE:6275978 5', mRNA sequence.		
ACCESSION	BQ675098		
VERSION	BQ675098.1	GI:21785932	
KEYWORDS	EST.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (bases 1 to 930)		
TITLE	NIH-MGC http://mgc.nci.nih.gov/.		
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)		
COMMENT	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: c9apbs-remail.nih.gov		
	Tissue Procurement: ATCC		
	cDNA Library Preparation: Rubin Laboratory		
	cDNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)		
	DNA distribution by: Agencourt Bioscience Corporation		
	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov		
	Plate: LLCM2460 row: a column: 03		
	High quality sequence stop: 616.		
FEATURES	Location/Qualifiers		
source	1..930		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:6275978"		
	/tissue_type="epidermoid carcinoma, cell line"		
	/lab_host="DH10B (phage-resistant)"		
	/clone_lib="NIH_MGC_102"		
	/notes="Organ: salivary gland; Vector: pOTB7; Site_1: XhoI; Site 2: EcoRI; CDNA made by oligo-dr priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."		
ORIGIN			
Query Match	34.1%;	Score 843.4;	DB 5; Length 930;
Best Local Similarity	97.8%;	Pred. No. 2e-161;	
Matches	886;	Conservative 0;	Mismatches 17; Indels 3; Gaps 3;
Qy	484	TTTTTGATTTTATAGCCACAAATGCTGCATATTCATACTTTAAATTCCTAAAGAAATAATTTTT	543
Db	1	TTTTGATTTTATAGCCACAAATGCTGCATATTCATACTTTAAATTCCTAAAGAAATAATTTTT	60
Qy	544	AATGTTTAAACCGTGATAATGCAATAAATAGAAAAATGTGGTTTACAAAAATAAAACCGTC	603
Db	61	AATGTTTAAACCGTGATAATGCAATAAATAGAAAAATGTGGTTTACAAAAATAAAACCGTC	120
Qy	604	TTCACTAGTTTACCACCTGAAGTAAGATGTCCTGGTTGGAGCTAAGACCATCATTTGTG	663
Db	121	TTCACTAGTTTACCACCTGAAGTAAGATGTCCTGGTTGGAGCTAAGACCATCATTTGTG	180

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: L10M698 row: d column: 20
High quality sequence stop: 615.
Location/Qualifiers

FEATURES

source
1. .932
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6536612"
/tissue_type="epidermoid carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_101"
/note="Organ: lung; Vector: pOTB7; Site_1: EcoRI; Site_2:
XhoI; cDNA made by oligo-dr priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."

ORIGIN

Query Match 32.3%; Score 797.8; DB 5; Length 932;
Best Local Similarity 99.5%; Pred. No. 4e-152;
Matches 82; Conservative 0; Mismatches 2; Indels 2; Gaps 2;
QY 794 ACAACCTCAGAGCTCTGCTCAGTCAGTCACCTTTTGGTCACACATCCCATTTTAAAG 853
DB 1 ACAACCTCAGAGCTCTGCTCAGTCAGTCACCTTTTGGTCACACATCCCATTTTAAAG 60
QY 854 ATCTGACAGCTCCATACAGATCATGTCTCAAGCTTCAGTATGCTGCTTATTCACA 913
DB 61 ATCTGACAGCTCCATACAGATCATGTCTCAAGCTTCAGTATGCTGCTTATTCACA 120
QY 914 GCTATTCTTTGCTGCAACCTGATTGAAAATGTTTCAGAGATTAGGCTTGCACCCACCACT 973
DB 121 GCTATTCTTTGCTGCAACCTGATTGAAAATGTTTCAGAGATTAGGCTTGCACCCACCACT 180
QY 974 GTCATTAGATTAAATAAATCTTTTGGAGTCTTTGTCATCATGTTATCTCAAGTCTGTATCC 1033
DB 181 GTCATTAGATTAAATAAATCTTTTGGAGTCTTTGTCATCATGTTATCTCAAGTCTGTATCC 240
QY 1034 TGTGGTGTGCAATCCAGTGGAGCTTTAGTAGTACTCAGATCTCTCTTTGTTGGTGGCT 1093
DB 241 TGTGGTGTGCAATCCAGTGGAGCTTTAGTAGTACTCAGATCTCTCTTTGTTGGTGGCT 300
QY 1094 AGTATATTAAAGTAAACCTGCTGTATGCTCACCAGAAAGCAACAGAGCATGTCAGT 1153
DB 301 AGTATATTAAAGTAAACCTGCTGTATGCTCACCAGAAAGCAACAGAGCATGTCAGT 360
QY 1154 GCTTTGATCTGAGAGCTTTTGTCTTACAAATCTCAGAAATGCTGAAGGCCACATCAT 1213
DB 361 GCTTTGATCTGAGAGCTTTTGTCTTACAAATCTCAGAAATGCTGAAGGCCACATCAT 420
QY 1214 TTAGGAAGAGTTTAAATGTACCTTTTAAAGGTCAAGAGTTATAGATTCACCTGATTA 1273
DB 421 TTAGGAAGAGTTTAAATGTACCTTTTAAAGGTCAAGAGTTATAGATTCACCTGATTA 480
QY 1274 CCTGGGATCTCATTGAAATGTCAGAGTTCAATTAATGAGGCTATTACCTATCAAAAA 1333
DB 481 CCTGGGATCTCATTGAAATGTCAGAGTTCAATTAATGAGGCTATTACCTATCAAAAA 540
QY 1334 TCAACTGCCCTCAAGGTGGCACTCTTTTGTACAACTTTTATCCGAGACACTTCTGACACT 1393
DB 541 TCAACTGCCCTCAAGGTGGCACTCTTTTGTACAACTTTTATCCGAGACACTTCTGACACT 600
QY 1394 GGAGAGAGGAAGTGTGGTGTGAGTTATGGGGTTTCTCTTGAAAATGCAAGTCTTTGGACCA 1453
DB 601 GGAGAGAGGAAGTGTGGTGTGAGTTATGGGGTTTCTCTTGAAAATGCAAGTCTTTGGACCA 660

QY 1454 CTGCTTAACCTAGGAGGAGCTAATCAGTGACACAGTAGATCTTGTCTGTGCCAAAA 1513
DB 661 CTGCTTAACCTAGGAGGAGCTAATCAGTGACACAGTAGATCTTGTCTGTGCCAAAA 720
QY 1514 GTTATACAT-CCATCTTTGAAGCAGTTTCTCAATATGTCATCGTATTTATTTGCCATAGACAG 1572
DB 721 GTTATACATCCCATCTTTGAAGCAGTTTCTCCATATGTCATCGTATTTATTTGCCATAGACAG 780
QY 1573 -AATTGGAGTACTCTGATGGAAACCCCTGACTAAAATGACAGAA 1616
DB 781 AAATTGGAGTACTCTGATGGAAACCCCTGACTAAAATGACAGAA 825

RESULT 11

BG742334 810 bp mRNA linear EST 15-MAY-2001
602631495F1 NCI_CGAP_skn3 Homo sapiens cDNA clone IMAGE:4776791 5',
mRNA sequence.
BG742334
ACCESSION BG742334
VERSION BG742334.1 GI:14052987
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 810)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: James Cleaver, M.D.
cDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA
Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: L10M10629 row: n column: 24
High quality sequence stop: 807.
Location/Qualifiers

FEATURES

source
1. .810
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4776791"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NCI_CGAP_skn3"
/note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dr.
Average insert size 1.5kb. Library constructed by Life
Technologies. Note: this is a NCI_CGAP Library."

ORIGIN

Query Match 31.8%; Score 785.4; DB 4; Length 810;
Best Local Similarity 99.6%; Pred. No. 1.3e-149;
Matches 808; Conservative 0; Mismatches 1; Indels 2; Gaps 2;
QY 1333 ATCAACTGCCCTCAAA-GGTGGCACTCTTTTGTACAACTTTATCCGAGACACTTCTGACA 1391
DB 1 ATCAACTGCCCTCAAGGTGGCACTCTTTTGTACAACTTTATCCGAGACACTTCTGACA 60
QY 1392 CTGGAGAAGAACTGTGTGGTTCAGTTATGGGGTTTCTCTTGAANAATGACGCTTGGACC 1451
DB 61 CTGGAGAAGAACTGTGTGGTTCAGTTATGGGGTTTCTCTTGAANAATGACGCTTGGACC 120
QY 1452 AGCTGCTTAACTAGGAAGGAGCTAATCAGTGACACACCTGATCTTGTCTGTGCCAAA 1511
DB 121 AGCTGCTTAACTAGGAAGGAGCTAATCAGTGACACACCTGATCTTGTCTGTGCCAAA 180
QY 1512 AAGTTATACATCCATCTTTGAAGCAGTTTCTCAATATGATCGTATTTATTTGCCATAGACA 1571

Db 181 AAGTTATACATCCATCTTTGAAGCAGGTTTCTCAATATGATCGTATTATTATGGCCATAGACA 240
QY 1572 GAATTCGAGTGAATCTGATGGAACCCCTGACTGCTAAATATGACAGAAACAGAGCTATTGGAT 1631
Db 241 GAATTCGAGTGAATCTGATGGAACCCCTGACTGCTAAATATGACAGAAACAGAGCTATTGGAT 300
QY 1632 CCTAGGCTCAATATGCTCTAATAGTATGGAAGTGTGAAGATGTGCACCTGCAAAAT 1691
Db 301 CCTAGGCTCAATATGCTCTAATAGTATGGAAGTGTGAAGATGTGCACCTGCAAAAT 360
QY 1692 TTGGCTCCCAACATTTTTTTTCATCTTATTCCTAATGAAGCAACAATCTGCAGCTTCTCTC 1751
Db 361 TTGGCTCCCAACATTTTTTTTCATCTTATTCCTAATGAAGCAACAATCTGCAGCTTCTCTC 420
QY 1752 TCTGCAACAGAAATGACACTGCTGGGATGAGCTGAAGCTCAGTGTGACAGCGGCACTGC 1811
Db 421 TCTGCAACAGAAATGACACTGCTGGGATGAGCTGAAGCTCAGTGTGACAGCGGCACTGC 480
QY 1812 ATGCTCTGCGAGTTAACTCAAGGAACCATGGGCTTTGTTGGAGGTGGCTGTACTGAA 1871
Db 481 ATGCTCTGCGAGTTAACTCAAGGAACCATGGGCTTTGTTGGAGGTGGCTGTACTGAA 540
QY 1872 CTCAATTTGGCTGCATATATCAGACACAAGACTCACAAACGACCCAGAAAGCAATTTCTCAAAG 1931
Db 541 CTCAATTTGGCTGCATATATCAGACACAAGACTCACAAACGACCCAGAAAGCAATTTCTCAAAG 600
QY 1932 ATGATGAATGTACTCAACAGAACTTCAATTAATTTGCTGAAGCAATTTGCGAGTGCCTAG 1991
Db 601 ATGATGAATGTACTCAACAGAACTTCAATTAATTTGCTGAAGCAATTTGCGAGTGCCTAG 660
QY 1992 AATCTGTGTGGCTCTTTAGAACATGATGGAGGTGAAATTTCTCACTGACATCAAGTATG 2051
Db 661 AATCTGTGTGGCTCTTTAGAACATGATGGAGGTGAAATTTCTCACTGACATCAAGTATG 720
QY 2052 GACACCTTTGGTCAGTTTCAGGAGATTTCTCCCTGTGTGTCTAACTGGCCAGATTTGCTTT 2111
Db 721 GACACCTTTGGTCAGTTTCAGGAGATTTCTCCCTGTGTGTCTAACTGGCCAGATTTGCTTT 780
QY 2112 CACAGTGTGGCTGTGATATACAAATAGCCA 2142
Db 781 CACAGTGTGG-TGTGGATATACAAATAGCCA 810

RESULT 12

BM913546
LOCUS AGENCOURT_6610292 NIH_MGC_41 Homo sapiens cdna clone IMAGE:5475679
DEFINITION 5', mRNA sequence. 1044 bp mRNA linear EST 12-MAR-2002

ACCESSION BM913546
VERSION BM913546.1 GI:19363925

KEYWORDS EST.
SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 1044)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/DTF

CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LNL at:
<http://image.lnl.gov>

Plate: LLCMI990 row: o column: 08
High quality sequence stop: 679.

Location/Qualifiers
1. .1044

/organism="Homo sapiens"

FEATURES

source

/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5475679"
/tissue_type="amelanotic melanoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_41"
/note="Organ: skin; Vector: pOTB7; Site 1: XhoI; Site 2:
EcORI; cDNA made by oligo-dT priming. Directionally cloned
into EcORI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Library constructed by Ling Hong in the
Laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."

ORIGIN

Query Match 31.6%; Score 781; DB 5; Length 1044;
Best Local Similarity 98.5%; Pred. No. 1.1e-148;
Matches 798; Conservative 0; Mismatches 11; Indels 1; Gaps 1;
QY 1612 AGGAACACAGCCTATTGGATCCCTAGGCTCAATATGCTCTAATAGTATGGAAGTGA 1671
Db 1 AGGAACACACGCTATTGGATCCCTAGGCTCAATATGCTCTAATAGTATGGAAGTGA 60
QY 1672 AGATGTGTGCACCTGCAAAATTTGGCTCCCAACATTTTTTTTCATCTTATTCCTAATGAAGC 1731
Db 61 AGATGTGTGCACCTGCAAAATTTGGCTCCCAACATTTTTTTTCATCTTATTCCTAATGAAGC 120
QY 1732 AACAAATCTCAGCTTGTCTTCTGCAACAGAAATGACACTGGCTGGGATGAGCTGAAGCT 1791
Db 121 AACAAATCTCAGCTTGTCTTCTGCAACAGAAATGACACTGGCTGGGATGAGCTGAAGCT 180
QY 1792 CAGCTGTGAGCGGCACTGCATGTCTGCAGTTAACTCAAGGAACCATGGCTTTGTT 1851
Db 181 CACGTGTGAGCGGCACTGCATGTCTGCAGTTAACTCAAGGAACCATGGCTTTGTT 240
QY 1852 GGGAGTGGCTGTACTGAACTCATTTGGCTGCATATATCAGACACAAGACTCACAACGA 1911
Db 241 GGGAGTGGCTGTACTGAACTCATTTGGCTGCATATATCAGACACAAGACTCACAACGA 300
QY 1912 CCCAGAAAGCAATTTCTCAAAAGATGATGAATGTACTCAACAGAACTTCAATTAATTCCTGA 1971
Db 301 CCCAGAAAGCAATTTCTCAAAAGATGATGAATGTACTCAACAGAACTTCAATTAATTCCTGA 360
QY 1972 AGCAATTTTGCAGTGCCTTAGAATCTGTTGGCTCTTTAGAACATGATGGAGTGA 2031
Db 361 AGCAATTTTGCAGTGCCTTAGAATCTGTTGGCTCTTTAGAACATGATGGAGTGA 420
QY 2032 TCTCACTGACATGAAGTATGGACACCTTTGGTCAGTTTCAGGAGATTTCTCCCTGTGTTGC 2091
Db 421 TCTCACTGACATGAAGTATGGACACCTTTGGTCAGTTTCAGGAGATTTCTCCCTGTGTTGC 480
QY 2092 TAACTGGCCAGATTTGCTTTTCAAGTGTGGCTGTGGAATATACAATAGCAGGAAGAACT 2151
Db 481 TAACTGGCCAGATTTGCTTTTCAAGTGTGGCTGTGGAATATACAATAGCAGGAAGAACT 540
QY 2152 CAACTGGCTTTTCTTAAGAGACACAGCTCGTCCATTTGTGCCCAAGCTGCCCTCCACA 2211
Db 541 CAACTGGCTTTTCTTAAGAGAGCACATGTCTCCATTTGTGCCCAAGCTGCCCTCCACA 600
QY 2212 TGAAGCTGTGGGCTCAGCCAGCAACCTGACCTTTGGACTGTTGACCTGCAAGCTTAGTGG 2271
Db 601 TGAAGCTGTGGTCTCAGCCAGCAACCTGACCTTTGGACTGTTGACCTGCAAGCTTAGTGG 660
QY 2272 CCTACAGGTGGCTGTAGACAGCCAAATTTGATTTGGGATCTTTTCAATATGTTATTGAAGA 2331
Db 661 CCTACAGGTGGCTGTAGACAGCCAAATTTGATTTGGATCTTTTCAATATGTTATTGAAGA 720
QY 2332 TAAAACTAAGAGATAGCATGTTTGGTATATACAGAGAACAA-ATAAACTAGTCTGTTG 2390
Db 721 TAAAACTAAGAGATAGCATGTTTGGTATATACAGAGAACCAANNNTAACTAGTCTGTTG 780
QY 2391 GCAATTTGAGAAAAA 2420

Db 781 GGCAATTGGAGAAAAA... 810

RESULT 13
B0423217
LOCUS
DEFINITION
AGENCOURT_7828752 NIH MGC_68 Homo sapiens cDNA clone IMAGE:6056596
5', mRNA sequence.

ACCESSION
VERSION
B0423217
KEYWORDS
EST.
SOURCE
Homo sapiens (human)

ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 912)
NIH-MGC <http://mgs.nci.nih.gov/>
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/Drp/Gazdar
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM1318 row: 1 column: 05
High quality sequence stop: 684.
Location/Qualifiers
1..912
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6056596"
/issue_type="large cell carcinoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 68"
/note="Organ: lung; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.8 kb. Library constructed by Life
Technologies."

ORIGIN
Query Match 30.9%; Score 765; DB 5; Length 912;
Best Local Similarity 98.6%; Pred. No. 1.9e-145;
Matches 802; Conservative 0; Mismatches 7; Indels 4; Gaps 3;
QY 1 TATGAGCCTTCGGAACCTGTGAGAGACTACAAAGTTTGGTGTGTTAGTCCCTTAGT 60
DB 100 TATGAGCCTTCGGAACCTGTGAGAGACTACAAAGTTTGGTGTGTTAGTCCCTTAGT 159
QY 61 TGGGCTCATACATTTGGGGTGGTACAGAAATCAAAAGCAGCCCTGTTTCCAAATACCTAA 120
DB 160 TGGGCTCATACATTTGGGGTGGTACAGAAATCAAAAGCAGCCCTGTTTCCAAATACCTAA 219
QY 121 AAACGACGACATCTCTGACGACAGATAGTCTGGGACTTTCAAATCTTCAGAGAGCCAAAT 180
DB 220 AAACGACGACATCTCTGACGACAGATAGTCTGGGACTTTCAAATCTTCAGAGAGCCAAAT 279
QY 181 CCAGGGGAAGTAGCAGGCTTCGAACTCTTCAGTAAAGAGCAGCTTTGAATCTGAGCTTC 240
DB 280 CCAGGGGAAGTAGCAGGCTTCGAACTCTTCAGTAAAGAGCAGCTTTGAATCTGAGCTTC 339
QY 241 ATATCGAAAGAGAGATGAAATACCAAGTTGGATTAGAAAGAACTGGCTTCTTTAGCT 300
DB 340 ATATCGAAAGAGAGATGAAATACCAAGTTGGATTAGAAAGAACTGGCTTCTTTAGCT 399
QY 301 GGGATATCTTCATAGGTGTCATCTTGGAAACATCTTTTGGAGAGGTCTGCAAGCAG 360
DB 400 GGGATATCTTCATAGGTGTCATCTTGGAAACATCTTTTGGAGAGGTCTGCAAGCAG 459

QY 361 TCTGTAAATTTTCAAGTCTCAAGCAAAACAAAAGAGTATTGAAGAGTGAAGTAAAT 420
DB 460 TCTGTAAATTTTCAAGTCTCAAGCAAAACAAAAGAGTATTGAAGAGTGAAGTAAAT 519
QY 421 ATTGGAAATTTCAATTTTGTCTCAATTAATCAATTTCTATGCTGATTAAGCAATTTGA 480
DB 520 ATTGGAAATTTCAATTTTGTCTCAATTAATCAATTTCTATGCTGATTAAGCAATTTGA 579
QY 481 ACTTTTTCATTTTATAGCCCAATGCTGCAATTTTATACATTTTAAATCTTCAAAAGTAAAT 540
DB 580 ACTTTTTCATTTTATAGCCCAATGCTGCAATTTTATACATTTTAAATCTTCAAAAGTAAAT 639
QY 541 TTTAATGTTTAAAGCGTGATATGCAATTAATAGAAAATGTTGTTTCAAAATTAAGCAAT 600
DB 640 TTTAATGTTTAAAGCGTGATATGCAATTAATAGAAAATGTTGTTTCAAAATTAAGCAAT 699
QY 601 GTCTTCACTAGTTACCACTGAGTGAAGTATGCTCTGTTTGGAGAGTCAAGAGCCATCAT 660
DB 700 GTCTTCACTAGTTACCACTGAGTGAAGTATGCTCTGTTTGGAGAGTCAAGAGCCATCAT 759
QY 661 GTGTAAGAGTCAACCACTGACCACTGAGAGAGTCAAGAGCCATCATCTTCTGTTGAAAG 720
DB 760 GTGTAAGAGTCAACCACTGACCACTGAGAGAGTCAAGAGCCATCATCTTCTGTTGAAAG 819
QY 721 AATTGTTAATCATGCTATGCCCCCTC-AGTGTAGGCTGAAGCAGCTGCACAAATGGCTTT- 778
DB 820 AATTGTTAATCATGCTATGCCCCCTC-AGTGTAGGCTGAAGCAGCTGCACAAATGGCTTT 879
QY 779 -GGAGTTTACGTGTGTA-CAACCTCAGATCCT 809
DB 880 GGGAGTTTACGTGTGTA-CAACCTCAGATCCT 912

RESULT 14
B0433357/c
LOCUS
DEFINITION
B0433357 Homo sapiens FETAL LIVER Homo sapiens cDNA clone
CS0DM014YF11 3-PRIME, mRNA sequence.

ACCESSION
VERSION
B0433357.2 GI:47002502
KEYWORDS
EST.
SOURCE
Homo sapiens (human)

ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 942)
Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
Full-length cDNA libraries and normalization
Unpublished (2001)
On May 15, 2003 this sequence version replaced gi:30775192.
Contact: Genoscope
Genoscope - Centre National de Sequencage
2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
Email: segre@genoscope.cns.fr, web : www.genoscope.cns.fr
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library
was not normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
This sequence belongs to sequence cluster 9906.r
For more information about this cluster, see
http://www.genoscope.cns.fr/cdna?s=CS0BA1052ZC11_CS04979_1&c=9906.r

FEATURES
source
1..942
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DM014YF11"
/tissue_type="FETAL LIVER"
/dev_stage="fetal"
/clone_lib="Homo sapiens FETAL LIVER"

/note="Organ: liver; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."

ORIGIN		Query Match	30.4%;	Score	752.2;	DB	5;	Length	942;
		Best Local Similarity	92.1%;	Pred. No.	8e-143;				
		Matches	856;	Conservative	0;	Mismatches	64;	Indels	9;
		Gaps	6;						
QY	940	AAATGTTACAGAGATTAGGCTTTGACACCCACCACTGTCATTAGATTAAATAAATCATCTTTT	999						
DB	922	AAATGTTCAAAATAGGCTTGACCAACCACTGTCAT-AGATTAAATAAATCATCTTT-	865						
QY	1000	GAGCTTTGCATCAGTTATCAAGTCTGATGCTGATGCTGTTGTCGAATCCAGTGACGTT	1059						
DB	864	--GAGCTTGCATCAGTATTTCAGTCTGAGACCTGGGGTT-TCGAATCCAGTGACGTT	808						
QY	1060	TAGTAGTACTCAGATCCTCTTTGTTTGGTGGCTAGTATATTA-ACAAGTAAACCTGCCT	1118						
DB	807	TAGTAGACTCAAGATCCTCCTTTGTTGGTGGCTAGTATATAACAAGTNNAAACCTGCCT	748						
QY	1119	GTATGCTCACCAAGAAAGCAAGACATGTCAGTCTTTGATCCTGAGAGCCTTTTTCG	1178						
DB	747	GTATGCTCACCAAGAAAGCAAGACATGTCAGTCTTTGATCCTGAGAG-CTTTTTCG	689						
QY	1179	TTCAAATTCAGAAAATGCTGAAGGCCACATCATTTTAGAAAGAGTTTAATGTACCTT	1238						
DB	688	TTCAAATTCAGAAAATGCTGAAGGCCACATCATTTTAGANAGAGTTTA--TTGTACTT	631						
QY	1239	TAAAGGTCAAAGATTATAGATTCCACTGTATTAATCTGGGATCTCATTTGAATGTCAG	1298						
DB	630	TAAAGGTCAAAGATTATAGATTCCACTGTATTAATCTGGGATCTCATTTGAATGTCAG	571						
QY	1299	AAGTTCAATTAATGAGGCTATTACCTATCAAAAATCAACTGCCTCCCAAGGTGGCACTCT	1358						
DB	570	AAGTTCAATTAATGAGGCTATTACCTATCAAAAATCAACTGCCTCCCAAGGTGGCACTCT	511						
QY	1359	TTTGTAACAATTTATCCGGAGACACTTCTGACACTGGAGAAGAACTGTGGTGGTCAGTT	1418						
DB	510	TTTGTAACAATTTATCCGGAGACACTTCTGACACTGGAGAAGAACTGTGGTGGTCAGTT	451						
QY	1419	ATGGGTTTCTCTTGAAAATGCACTCTGACACAGCTGCTTAACCTAGGAGGCACCTAA	1478						
DB	450	ATGGGTTTCTCTTGAAAATGCACTCTGACACAGCTGCTTAACCTAGGAGGCACCTAA	391						
QY	1479	TCAGTGACCACTAGATCTTGCTGTGCCCCAAAAGTTATACATCCATCTTTTGAAGCAGT	1538						
DB	390	TCAGTGACCACTAGATCTTGCTGTGCCCCAAAAGTTATACATCCATCTTTTGAAGCAGT	331						
QY	1539	TTCTCAATATGCATTCGATTATTATGCCATAGACAGAAATGGAGTGACTCTGATGGAACCCC	1598						
DB	330	TTCTCAATATGCATTCGATTATTATGCCATAGACAGAAATGGAGTGACTCTGATGGAACCCC	271						
QY	1599	TGACTAAAATGACAGGACACAGCCTATTGGATCCCTAGGCTCAATATGTCCTAATAGTT	1658						
DB	270	TGACTAAAATGACAGGACACAGCCTATTGGATCCCTAGGCTCAATATGTCCTAATAGTT	211						
QY	1659	ATGGAAGTGGAAGATGTGCACTGCAAAAATTTGGCTCCCAACATTTTTTTCATCTTA	1718						
DB	210	ATGGAAGTGGAAGATGTGCACTGCAAAAATTTGGCTCCCAACATTTTTTTCATCTTA	151						
QY	1719	TTCTTAATGAAGCAACAATCTGACGCTTGCTTCTCTGCAACAGAAATGACATGCTCGTGG	1778						
DB	150	TTCTTAATGAAGCAACAATCTGACGCTTGCTTCTCTGCAACAGAAATGACATGCTCGTGG	91						
QY	1779	ATGAGCTGAAGCTCAGCTGTGACGCGCACTGATGCTCTGCGATTAACTCAAGGAAC	1838						
DB	90	ATGAGCTGAAGCTCAGCTGTGACGCGCACTGATGCTCTGCGATTAACTCAAGGAAC	31						
QY	1839	CATGGGCTTTGTTGGGAGGTGGCTGTACT	1867						

Db	30	CATGGCTTGTGGAGGGGCTGACTAACT	2						
RESULT 15		AL532425/c							
LOCUS		AL532425							
DEFINITION		AL532425 Homo sapiens FETAL LIVER Homo sapiens cDNA clone							
ACCESSION		AL532425							
VERSION		AL532425.3	GI:45707342						
KEYWORDS		EST.							
SOURCE		Homo sapiens (human)							
ORGANISM		Homo sapiens							
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.							
AUTHORS		Li, W.B., Gruber, C., Jessee, J. and Polayes, D.							
TITLE		Full-length cDNA libraries and normalization							
JOURNAL		Unpublished (2001)							
COMMENT		On Feb 13, 2001 this sequence version replaced gi:31070257.							
		Contact: Genoscope							
		Genoscope - Centre National de Sequencage							
		2 rue Gaston Cremieux, Cp 5706 - 91057 EVRY cedex - FRANCE							
		Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr							
		1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized. Library was constructed by Life Technologies, a division of Invitrogen.							
		This sequence belongs to sequence cluster 9906.r							
		For more information about this cluster, see							
		http://www.genoscope.cns.fr/cdna?s=CS0DM014CC06NP16c=9906.r.							
FEATURES		Location/Qualifiers							
		1..1013							
		/organism="Homo sapiens"							
		/mol_type="mRNA"							
		/db_xref="taxon:9606"							
		/clone="CS0DM014YF11"							
		/tissue_type="FETAL LIVER"							
		/dev_stage="fetal"							
		/clone_lib="Homo sapiens FETAL LIVER"							
		/note="Organ: liver; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."							
ORIGIN		Query Match	29.9%;	Score	738;	DB	1;	Length	1013;
		Best Local Similarity	90.1%;	Pred. No.	6.3e-140;				
		Matches	798;	Conservative	35;	Mismatches	48;	Indels	5;
		Gaps	5;						
QY	1432	TGAAATGACGCTTTGGACCACTGCTTAACCTAGGAAGCAGCTAATCAGTGACACCGT	1491						
DB	938	TGAAATGACGCTT-GACCMGCTGCTT-ACCTAKGAGGCGAGCAGTACGTGMCACCT	881						
QY	1492	AGATCTTGCTGTGCCAAAAGTTATACATCCATCTTTGAAGCAGTTCTCA-ATATGC	1550						
DB	880	AGATCTTGCTGTGCCAAAAGTTATACATCCATCTATGAGCAGTTCTCAGATATGC	821						
QY	1551	ATCGTATTATTGCCATAGACAGAAATGGAGTGACTCTGATGGAACCCCTGACTTAAATGCA	1610						
DB	820	ATCGTATTATTGCCATAGACAGAAATGGAGTGACTCTGATGGAACCCCTGACTTAAATGCA	761						
QY	1611	CAGGAACACAGCCTATTGGATCCCTAGGCTCAATATGCTCTAATAGTTATGGAAGTGTGA	1670						
DB	760	CAGGAACACAGCCTATTGGATCCCTAGGCTCAATATGCTCTAATAGTTATGGAAGTGTGA	702						
QY	1671	AGATGTGTGACTGCAAAATTTGGCTCCCAACATTTTTTTCATCTTATCTTAATGAAG	1730						
DB	701	AAGAAGTGTGAACCTKCCCTATGGCTCCCACTCTCTTTCTCTCTCTCTCTCTCTCTCT	642						
QY	1731	CAACATCTGCAGCTTGCTTCTCTCTGCAACAGAAATGACACTGCCTGGGATGAGCTGAAGC	1790						

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 6, 2005, 06:59:18 ; Search time 10581 Seconds
(without alignments)
11320.413 Million cell updates/sec

Title: US-10-616-263-29
Perfect score: 2472
Sequence: 1 tatgagccttcggaacttgt.....aaaaaaaaaaaaaaaaaaaa 2472

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0
Searched: 4708233 seqs, 24227607955 residues
Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl.*
1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2472	100.0	2472	6	BD227254 Secreted
2	2429.2	98.3	2497	9	BC028973 Homo sapi
3	2412.2	97.6	2678	9	AF221993 Homo sapi
4	2412	97.6	2512	9	AL157427 Homo sapi
5	2401	97.1	2516	9	AF221992 Homo sapi
6	2398.4	97.0	2503	9	AK025741 Homo sapi
7	1763.2	71.3	1921	6	BD276479 143 Human
8	1708.2	69.1	1713	6	CQ729071 Sequence
9	1571	63.6	2972	10	BC083863 Rattus no
10	1547.4	62.6	2967	10	BC080765 Mus muscu
11	1545.8	62.5	2699	10	AF254074 Mus muscu
12	1405.6	56.9	74813	9	AL034430 Human DNA
13	1253.8	50.7	1274	6	BD276497 143 Human
14	908.2	36.7	207304	2	AC107146 Rattus no
15	908.2	36.7	259468	2	AC096067 Rattus no
16	875.6	35.4	195301	10	AL731706 Mouse DNA
17	627	25.4	628	9	HUMY132F03 Homo sapi
18	587.4	23.8	679	9	AF275813 Homo sapi
19	577.8	23.4	1173	10	BC024359 Mus muscu

C	20	457.4	18.5	494	11	G60203	G60203 SHGC-131114
C	21	446.4	18.1	491	6	CQ690654	CQ690654 Sequence
C	22	422.4	17.1	458	6	BD139286	BD139286 Extended
C	23	413.4	16.7	421	6	BD265148	BD265148 Compounds
C	24	413.4	16.7	421	6	AR401134	AR401134 Sequence
C	25	413.4	16.7	421	6	AX192580	AX192580 Sequence
C	26	371.2	15.0	1831	5	BC045401	BC045401 Danio rer
C	27	344.6	13.9	348	6	AX887034	AX887034 Sequence
C	28	344.6	13.9	348	6	BD026644	BD026644 Sequence
C	29	327.8	13.3	371	6	BD058768	BD058768 Secreted
C	30	320.4	13.0	428	11	BV103260	BV103260 MARC 1718
C	31	311.2	12.6	956	10	BC058174	BC058174 Mus muscu
C	32	279.4	11.3	412	11	BV093483	BV093483 RPAMSEQ0
C	33	270.8	11.0	378	11	BV093401	BV093401 RPAMSEQ0
C	34	269	10.9	332	6	AX261806	AX261806 Sequence
C	35	266.2	10.8	387	11	BV093484	BV093484 RPAMSEQ0
C	36	263.2	10.6	375	11	BV093400	BV093400 RPAMSEQ0
C	37	262.4	10.6	390	11	BV093485	BV093485 RPAMSEQ0
C	38	262.4	10.6	390	11	BV159500	BV159500 RPAMSEQ0
C	39	260	10.5	379	11	BV093486	BV093486 RPAMSEQ0
C	40	260	10.5	379	11	BV157816	BV157816 RPAMSEQ0
C	41	247.6	10.0	1093	5	BX930888	BX930888 Gallus ga
C	42	246.4	10.0	1099	5	BX934020	BX934020 Gallus ga
C	43	214.6	8.7	295	11	BV093490	BV093490 RPAMSEQ0
C	44	214	8.7	291	11	BV157818	BV157818 RPAMSEQ0
C	45	205.8	8.3	385	11	BV093388	BV093388 RPAMSEQ0

ALIGNMENTS

RESULT 1	BD227254	2472 bp	DNA	linear	PAT 17-JUL-2003
LOCUS	Secreted proteins and polynucleotides encoding them.				
DEFINITION	BD227254				
ACCESSION	BD227254.1	GI:33037024			
VERSION	JP 2002522062-A/15.				
KEYWORDS	Homo sapiens (human)				
SOURCE	Homo sapiens				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 2472)				
AUTHORS	Jacobs, K., McCoy, J.M., Lavallie, E.R., Racie, L.A.C., Evans, C., Merberg, D., Treacy, M., Agostino, M.J., Ii, R.J.S., Spaulding, V., Wong, G.G., Clark, H.F. and Fechtel, K.				
TITLE	Secreted proteins and polynucleotides encoding them				
JOURNAL	Patent: JP 2002522062-A 15 23-JUL-2002; GENETICS INSTITUTE INC				
COMMENT	OS Homo sapiens (human) PN JP 2002522062-A/15 PD 23-JUL-2002 PF 13-AUG-1999 JP 2000565001 PR 14-AUG-1998 US 60/096622,17-AUG-1998 US 60/096815 PR 04-SEP-1998 US 60/093229,23-OCT-1998 US 60/105368 PR 08-JAN-1999 US 60/115234,12-FEB-1999 US 60/119931 PR 18-FEB-1999 US 60/120575,30-APR-1999 US 60/132020 PR 11-AUG-1999 US 60/148424 PI KENNETH JACOBS, JOHN M MCCOY, EDWARD R LAVALLIE, LISA A COLLINS PI RACIE, PI CHERYL EVANS, DAVID MERBERG, MAURICE TREACY, MICHAEL J AGOSTINO, PI ROBERT J STEININGER II, VIKKI SPAULDING, GORDON G WONG, HILARY F CLARK, PI KIM FECHTEL PI KIM FECHTEL PC C12N15/09, A61K38/00, A61K39/395, A61K45/00, A61K48/00, A61P7/00, PC A61P7/02, A61P7/04, A61P7/06, A61P13/00, A61P29/00, A61P35/00, PC A61P37/02, PC A61P43/00, A61P43/00, C07K14/47, C12N5/10, C12P21/02, G01N33/15, PC G01N33/50, PC C12N15/00, C12N5/00, A61K37/02 CC Secreted proteins and polynucleotides encoding them FH Key Location/Qualifiers				

FEATURES		FT	source	1. .2472	/organism='Homo sapiens (human)'	Location/Qualifiers	1. .2472
source		FT	source	1. .2472	/organism="Homo sapiens"	/mol_type="genomic DNA"	/db_xref="taxon:9606"
ORIGIN							
Query Match				100.0%;	Score 2472;	DB 6;	Length 2472;
Best Local Similarity				100.0%;	Pred. No. 0;		
Matches 2472;				Conservative	0;	Mismatches	0; Indels 0; Gaps 0;
Qy	1	TATGAGCCCTCGGAACCTGTGGAGAGACTACAAAGTTTGGTGTGTTATGTCCTT	60				
Db	1	TATGAGCCCTCGGAACCTGTGGAGAGACTACAAAGTTTGGTGTGTTATGTCCTT	60				
Qy	61	TGGGCTCATACATTTGGGGTGGTACAGAATCAAAAGCAGCCCTGTTTCCAAATACCTAA	120				
Db	61	TGGGCTCATACATTTGGGGTGGTACAGAATCAAAAGCAGCCCTGTTTCCAAATACCTAA	120				
Qy	121	AAACGACGACATTCCTGAGCAAGATAGTCTGGGACTTTTCAAAATCTTCAGAAGAGCCAAAT	180				
Db	121	AAACGACGACATTCCTGAGCAAGATAGTCTGGGACTTTTCAAAATCTTCAGAAGAGCCAAAT	180				
Qy	181	CNAGGGGAAGTAGCAGGCTTGCATCTTCAGGTAAAGAGCAGCTTTGAAATCTGAGCTTC	240				
Db	181	CNAGGGGAAGTAGCAGGCTTGCATCTTCAGGTAAAGAGCAGCTTTGAAATCTGAGCTTC	240				
Qy	241	ATATCGAAAGAGAGATGAAATATACAGTTGAGTTAGAAAGAACTGGCTTCTTGTAGCT	300				
Db	241	ATATCGAAAGAGAGATGAAATATACAGTTGAGTTAGAAAGAACTGGCTTCTTGTAGCT	300				
Qy	301	GGGATATCTTTCATAGGTGTCCATCTTTGGAACATACCTTTTTCAGAGGTCCTGCAAGCAG	360				
Db	301	GGGATATCTTTCATAGGTGTCCATCTTTGGAACATACCTTTTTCAGAGGTCCTGCAAGCAG	360				
Qy	361	TCGTAAATTTTCAGTCTCAAGCAAAACAAAGAGATTTGAAGAGTGAATTAATTAAT	420				
Db	361	TCGTAAATTTTCAGTCTCAAGCAAAACAAAGAGATTTGAAGAGTGAATTAATTAAT	420				
Qy	421	ATTGGAAATACATAATTTGTCTAATAATCATCTATGCTGATTAGCTTCATAAAACATTGA	480				
Db	421	ATTGGAAATACATAATTTGTCTAATAATCATCTATGCTGATTAGCTTCATAAAACATTGA	480				
Qy	481	ACTTTTGTATTTATAGCCACAAATGCTGCATATTCATCTTTAAATTCCTAAAGATAAT	540				
Db	481	ACTTTTGTATTTATAGCCACAAATGCTGCATATTCATCTTTAAATTCCTAAAGATAAT	540				
Qy	541	TTTAATGTTAAACGTTGAATGCAATAATAGAAAATGTGGTTTACAAAATAAACCG	600				
Db	541	TTTAATGTTAAACGTTGAATGCAATAATAGAAAATGTGGTTTACAAAATAAACCG	600				
Qy	601	GTCTTCACATAGTTTACCACTGAAGTAAAGATGTCTCGTTTGGAGCTAAGAAGCCATCAT	660				
Db	601	GTCTTCACATAGTTTACCACTGAAGTAAAGATGTCTCGTTTGGAGCTAAGAAGCCATCAT	660				
Qy	661	GTGTAAGAGTGAAACCACTGACAACTGAGAGGTCAGGACCAACATTTTCTGTTGAAAAG	720				
Db	661	GTGTAAGAGTGAAACCACTGACAACTGAGAGGTCAGGACCAACATTTTCTGTTGAAAAG	720				
Qy	721	AATTGTAAACATCATGCTATGGCCCTCAGGTAGGCTGAAGCAGCTGCACAATGGCTTTGG	780				
Db	721	AATTGTAAACATCATGCTATGGCCCTCAGGTAGGCTGAAGCAGCTGCACAATGGCTTTGG	780				
Qy	781	AGGTTACGTTGTGACAACTCTCAGTCTCAGCTCTGCTCAGTACACTTTTGGTACACA	840				
Db	781	AGGTTACGTTGTGACAACTCTCAGTCTCAGCTCTGCTCAGTACACTTTTGGTACACA	840				
Qy	841	TCCCATTTTAAAGATCCTGACAGCCCTCCATFACAGAATCATGTGTCAAGCTTCAGTGATTG	900				
Db	841	TCCCATTTTAAAGATCCTGACAGCCCTCCATFACAGAATCATGTGTCAAGCTTCAGTGATTG	900				

Qy	901	TGGCTTATTTCAGCTATTCTTTGCTGCAACCTGATTGAAATGTTTCAGAGATTAGGCTT	960
Db	901	TGGCTTATTTCAGCTATTCTTTGCTGCAACCTGATTGAAATGTTTCAGAGATTAGGCTT	960
Qy	961	GACACCCACCCTGTCTATTAGATTAAATAAAACATCTTTTGTAGCTTTTGTGATCTATCT	1020
Db	961	GACACCCACCCTGTCTATTAGATTAAATAAAACATCTTTTGTAGCTTTTGTGATCTATCT	1020
Qy	1021	CAAGTCTGATACCTGTGGTTCGAATCCAGTGAGCTTTTAGTAGTACTTCAGATCTCTCT	1080
Db	1021	CAAGTCTGATACCTGTGGTTCGAATCCAGTGAGCTTTTAGTAGTACTTCAGATCTCTCT	1080
Qy	1081	TTGTTTGGTGGCTAGTATATTAAACAAGTAAACCTGCTGTATGCTCACACAGAAAGAAAC	1140
Db	1081	TTGTTTGGTGGCTAGTATATTAAACAAGTAAACCTGCTGTATGCTCACACAGAAAGAAAC	1140
Qy	1141	AGAGCATGTCTAGTCTTTGATCCTCGAGAGCCCTTTTGTCTTACAATTCAGAAAAATGCTGA	1200
Db	1141	AGAGCATGTCTAGTCTTTGATCCTCGAGAGCCCTTTTGTCTTACAATTCAGAAAAATGCTGA	1200
Qy	1201	AGGCCACATCATTTTGAAGAGATTTAATTGTACTCTTTTAAAGGTCAAAGATTATAGA	1260
Db	1201	AGGCCACATCATTTTGAAGAGATTTAATTGTACTCTTTTAAAGGTCAAAGATTATAGA	1260
Qy	1261	TTCCCACTGTATTACCTGGGATACTCATTTGAAATGTGAGAAGTTCAATTAAATCAGGCTAT	1320
Db	1261	TTCCCACTGTATTACCTGGGATACTCATTTGAAATGTGAGAAGTTCAATTAAATCAGGCTAT	1320
Qy	1321	ACCTATCAAAAAATCAACTGCCCTCAAGGTGGCACTCTTTTGTACAACCTTTATCCGGAGA	1380
Db	1321	ACCTATCAAAAAATCAACTGCCCTCAAGGTGGCACTCTTTTGTACAACCTTTATCCGGAGA	1380
Qy	1381	CACCTCTGACACATGGAGAAGGAACCTGTGGTGTGCTAGTTATGGGGTTTCTCTTGAATAATGC	1440
Db	1381	CACCTCTGACACATGGAGAAGGAACCTGTGGTGTGCTAGTTATGGGGTTTCTCTTGAATAATGC	1440
Qy	1441	AGCTTTGGACCACTGCTTAACTAGGAAGGCGAGCTTAATCAGTACCACTGATGATCTCT	1500
Db	1441	AGCTTTGGACCACTGCTTAACTAGGAAGGCGAGCTTAATCAGTACCACTGATGATCTCT	1500
Qy	1501	CCTGTGCCAAAAAGTTATACATCCATCTTTTGAAGCAGTTTCTCAATATGCAATCGTATTAT	1560
Db	1501	CCTGTGCCAAAAAGTTATACATCCATCTTTTGAAGCAGTTTCTCAATATGCAATCGTATTAT	1560
Qy	1561	TGCCATAGACAGAAATTTGGAGTACTCTGATGGAACCCCTGACTTAAATGACAGGACACA	1620
Db	1561	TGCCATAGACAGAAATTTGGAGTACTCTGATGGAACCCCTGACTTAAATGACAGGACACA	1620
Qy	1621	GCCTATTGGATCCCTAGGCTCAATATGCTTAATAGTATGCTTAATGGAAGATGTGTG	1680
Db	1621	GCCTATTGGATCCCTAGGCTCAATATGCTTAATAGTATGCTTAATGGAAGATGTGTG	1680
Qy	1681	CAGTGCAAAAATTTGGCTCCCAACATTTTTCATCTTATTCCTAATGAAGCAACAATCTG	1740
Db	1681	CAGTGCAAAAATTTGGCTCCCAACATTTTTCATCTTATTCCTAATGAAGCAACAATCTG	1740
Qy	1741	CAGCTTGTCTTCTCTGCAACAGAAATGACACTGCTGCTGGGATGAGCTCAAGCTCACTG	1800
Db	1741	CAGCTTGTCTTCTCTGCAACAGAAATGACACTGCTGCTGGGATGAGCTCAAGCTCACTG	1800
Qy	1801	GACGCACTGCAATGCTGCACTTAACACTCAAGGAAACCATGGCTTTTCTGGAGGTGG	1860
Db	1801	GACGCACTGCAATGCTGCACTTAACACTCAAGGAAACCATGGCTTTTCTGGAGGTGG	1860
Qy	1861	CTGTACTGAAACTCAATTTGGCTGCATATATACAGACTCAAGACTCAACAGCCCAAGAAAG	1920
Db	1861	CTGTACTGAAACTCAATTTGGCTGCATATATACAGACTCAAGACTCAACAGCCCAAGAAAG	1920
Qy	1921	CAATCTCAAGATGATGATGATCTCAAAACAGAACTTCAATTAATGCTGAAGCATTTTG	1980
Db	1921	CAATCTCAAGATGATGATGATCTCAAAACAGAACTTCAATTAATGCTGAAGCATTTTG	1980
Qy	1981	CAGTGCCTTGAATCTGTGTTGGCTCTTTTAGAAACATGATGAGGTGAAATTTCTCACTGA	2040

Db	1981	CAGTGCCTAGAAATCTGTGTGGCTCTTTAGAACATGATGAGGTTGAAATTTCTCACTGA	2040	
Qy	2041	CATGAAGTATGGACACCTTTGTCAGTTTCAGTTCAGGAGATTTCTCCCTGTGTTCTCACTAGCGCC	2100	
Db	2041	CATGAAGTATGGACACCTTTGTCAGTTTCAGTTCAGGAGATTTCTCCCTGTGTTCTCACTAGCGCC	2100	
Qy	2101	AGATTTGCTTTTCACAGTGTGGCTGTGATTTATACAAATAGCCAGAGAACTCAACTGGTC	2160	
Db	2101	AGATTTGCTTTTCACAGTGTGGCTGTGATTTATACAAATAGCCAGAGAACTCAACTGGTC	2160	
Qy	2161	TTTCTTAAGAGACACGTCGTCAATTTGTCGCAAAAGCTGCTTCCCATGAAGCTGT	2220	
Db	2161	TTTCTTAAGAGACACGTCGTCAATTTGTCGCAAAAGCTGCTTCCCATGAAGCTGT	2220	
Qy	2221	GGGCTCAGCAGCAACCTTGACCTTGGACTGTTTGGACTGCAAGCTTAGTGGCTCAGAGT	2280	
Db	2221	GGGCTCAGCAGCAACCTTGACCTTGGACTGTTTGGACTGCAAGCTTAGTGGCTCAGAGT	2280	
Qy	2281	GGCTGTAGACAGCCCAATTTGATTTGGGATCTTTTCATATGTTATTGAAGATAAAACTA	2340	
Db	2281	GGCTGTAGACAGCCCAATTTGATTTGGGATCTTTTCATATGTTATTGAAGATAAAACTA	2340	
Qy	2341	AGAGAATAGCATGTTCTGTTATACAGAGAAACAAATAAATAGTCTGTTGGCAATTGAGA	2400	
Db	2341	AGAGAATAGCATGTTCTGTTATACAGAGAAACAAATAAATAGTCTGTTGGCAATTGAGA	2400	
Qy	2401	AA	2460	
Db	2401	AA	2460	
Qy	2461	AAAAAAAAAAAAA 2472		
Db	2461	AAAAAAAAAAAAA 2472		
RESULT 2	BC028973	2497 bp mRNA linear PRI 30-JUN-2004		
LOCUS	Homo sapiens McKusick-Kaufman syndrome, transcript variant 2, mRNA			
DEFINITION	(CDNA clone MGC:29831 IMAGE:3921348), complete cds.			
ACCESSION	BC028973			
VERSION	BC028973.2 GI:34783542			
KEYWORDS	MGC.			
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
REFERENCE	1 (bases 1 to 2497)			
AUTHORS	Klausner, R.D., Collins, F.S., Wagner, L., Shennan, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.P., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullah, S.J., Bosak, S.A., McEwan, P.J., McKernan, R.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmitz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalusz, D.E., Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.			
TITLE	Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences			
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)			
PUBMED	12477932			
REFERENCE	2 (bases 1 to 2497)			
AUTHORS	Strausberg, R.			
TITLE	Direct Submission			
JOURNAL	Submitted (01-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA			
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov			
COMMENT	On Sep 16, 2003 this sequence version replaced gi:23468371. Contact: MGC help desk Email: cgapbs-r@mail.nih.gov Tissue Procurement: ATCC/DCTD/DTF CDNA Library Preparation: Life Technologies, Inc. DNA Sequencing Arrayed by: The I.M.A.G.E. Consortium (LLNL) Sequencing Center: Baylor College of Medicine Human Genome Sequencing Center Center code: BCM-HGSC Web site: http://www.hgsc.bcm.tmc.edu/cdna/ Contact: amg@bcm.tmc.edu Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Lounseged, H., Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati, A.N., Gibbs, R.A.			
FEATURES	Location/Qualifiers			
source	1..2497 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="MGC:29831 IMAGE:3921348" /tissue_type="Skin, melanotic melanoma." /clone_lib="NIH_MGC_72" /lab_host="DH10B" /note="Vector: pCMV-SPORT6"			
gene	1..2497 /gene="MKKS" /note="synonyms: HMCS, KMS, MKS, BBS6" /db_xref="LocusID:8195" /db_xref="MIM:604896" 684..2396 /gene="MKKS" /codon_start=1 /product="McKusick-Kaufman syndrome protein"			
CDS	/protein_id="AAH28973.1" /db_xref="GI:23468372" /db_xref="LocusID:8195" /db_xref="MIM:604896"			
ORIGIN	translation="MSRLFAKKPSLCKSPLTTERVTTLSVLKRVITSCYGPSRLK QLNHFGGVCVTSQSSALLSHLVTHPLKLTASIQNHVSSPSDCGLFTAILCCNL IENVRLGUTPTVIRLNKHLSCISYLSKSTCGCRIPDFSSQILICLVRSILTS KFCAMLTRKRETHVSALILRAFLLTPENAEGHIIIGKSLIVPLKGQVIDSTVLPFI LTEMSEVQLMRLLPIPKKSTALKVALFCTTSGDSTDTGEGTVVSVYSLNVLQPL LNLQRLISDHVDLVLCOKIHPKSLQFLNMHRIIAIDRIGVTLMDEPLTKMTQPIG SLGSIICPNSYSGVKQVCTAKFGSKHFEHLI PNEATICSLLCNRNDANDELKLTQQT ALHVLQTLKEPALLGGCTETHLAAVIRHKTHNDPESILKDDDECTQTELQILAEAF CSALSSVSGSLEHGDGEIUTDMKYGHLMSVQADSPCVANWPDLLSCGGCLNYSBEL NWSFLRSTCRPFVQSCLPHEAVVSASNLTLDCLTAKUSGLQVAVETANLILDLSTVI EDKN"			
Query Match	98.3%;	Score 2429.2;	DB 9;	Length 2497;
Best Local Similarity	99.7%;	Pred. No. 0;		
Matches 2434;	Conservative 0;	Mismatches 8;	Indels 0;	Gaps 0;
Qy	1	TATGAGCCCTCGGAACCTGTGAGAGACTACAAAGTTTGGTGTGTTATGTCCTCTTAGT	60	
Db	56	TATGAGCCCTCGGAACCTGTGAGAGACTACAAAGTTTGGTGTGTTATGTCCTCTTAGT	115	
Qy	61	TGGGCTCATACATTTGGGCTGGTACAGAAATCAAAAGCAGCCCTGTTTCCAAATACCTAA	120	
Db	116	TGGGCTCATACATTTGGGCTGGTACAGAAATCAAAAGCAGCCCTGTTTCCAAATACCTAA	175	

QY	121	AAACGACGACATTCCTGACGACAGATAGTCTGGACCTTTCAATCTTTCAGAAAGCCCAAT	180
Db	176	AAACGACGACATTCCTGACGACAGATAGTCTGGACCTTTCAATCTTTCAGAAAGCCCAAT	235
QY	181	CCAGGGGAAGTAGCAGGCTTGCAATCTTCAGGTAAAGAACGACGCTTTGAATCTGAGCTTC	240
Db	236	CCAGGGGAAGTAGCAGGCTTGCAATCTTCAGGTAAAGAACGACGCTTTGAATCTGAGCTTC	295
QY	241	ATATCGAAAGAGAGATGAAATAATACACAGTTGGATTAGAAAGAACTGGCTTCCTGAGCT	300
Db	296	ATATCGAAAGAGAGATGAAATAATACACAGTTGGATTAGAAAGAACTGGCTTCCTGAGCT	355
QY	301	GGGATATCTTTCATAGGTGTCATCTTGGACATATCTTTTCGACAGGCTCTCGAAGCAG	360
Db	356	GGGATATCTTTCATAGGTGTCATCTTGGACATATCTTTTCGACAGGCTCTCGAAGCAG	415
QY	361	TCGTAAAAATTTTCAGTCTCAAGCAAAACAAAGAGATTTGAAGAGTGAAGTAAATAAAT	420
Db	416	TCGTAAAAATTTTCAGTCTCAAGCAAAACAAAGAGATTTGAAGAGTGAAGTAAATAAAT	475
QY	421	ATTTGGAAATTAATAATTTGTCAATAATCATTCATGCTGATTTAGCTTCATAAACATTGA	480
Db	476	ATTTGGAAATTAATAATTTGTCAATAATCATTCATGCTGATTTAGCTTCATAAACATTGA	535
QY	481	ACTTTTGTGATTTATAGCCACAAATGCTGCATATTCATCTTTAAATTCCTAAAGATAAT	540
Db	536	ACTTTTGTGATTTATAGCCACAAATGCTGCATATTCATCTTTAAATTCCTAAAGATAAT	595
QY	541	TTTAAATGTTTAAACGTTGATAATGCAATAAATAGAAAAATGTGGTTTACAAAAATAAAACG	600
Db	596	TTTAAATGTTTAAACGTTGATAATGCAATAAATAGAAAAATGTGGTTTACAAAAATAAAACG	655
QY	601	GTCTTCACTAGTTTACCACTGAGTAAAGATGCTCGTTTGGAGCTAAGAACCCATCAT	660
Db	656	GTCTTCACTAGTTTACCACTGAGTAAAGATGCTCGTTTGGAGCTAAGAACCCATCAT	715
QY	661	GTGTAGAGTGAACCACTGACACTGAGAGTGCAGGACCACTTTCTGTCTTGAAG	720
Db	716	GTGTAGAGTGAACCACTGACACTGAGAGTGCAGGACCACTTTCTGTCTTGAAG	775
QY	721	AATTTGAAACATCATGTATGGCCCTTCAGTAGGCTGAAAGCAGCTGCACAAATGGCTTTGG	780
Db	776	AATTTGAAACATCATGTATGGCCCTTCAGTAGGCTGAAAGCAGCTGCACAAATGGCTTTGG	835
QY	781	AGTTACGTTGTGTAACCTTCAAGTCTCAAGTCTGCTCAGTCACTTTTGGTCAACA	840
Db	836	AGTTACGTTGTGTAACCTTCAAGTCTCAAGTCTGCTCAGTCACTTTTGGTCAACA	895
QY	841	TCCCATTTTAAAGATCTGACAGCTTCCATACAGATCATGTGTCAAGCTTCAGTGATTG	900
Db	896	TCCCATTTTAAAGATCTGACAGCTTCCATACAGATCATGTGTCAAGCTTCAGTGATTG	955
QY	901	TGGCTTAATTCACAGCTATTCTTTGCTGCAACCTTGATTGAAAAATGTTTCAGAGATTAGGCTT	960
Db	956	TGGCTTAATTCACAGCTATTCTTTGCTGCAACCTTGATTGAAAAATGTTTCAGAGATTAGGCTT	1015
QY	961	GACACCAACATGTCATTTAGATTAAATAAACAATCTTTTGAAGTCTTTTGCATCAGTTATCT	1020
Db	1016	GACACCAACATGTCATTTAGATTAAATAAACAATCTTTTGAAGTCTTTTGCATCAGTTATCT	1075
QY	1021	CAAGTCTGATACCTGTGGTGTGCAATCCAGTGGACCTTTAGTAGTACTCAGATCTCTCT	1080
Db	1076	CAAGTCTGATACCTGTGGTGTGCAATCCAGTGGACCTTTAGTAGTACTCAGATCTCTCT	1135
QY	1081	TTGTTTGGTGGCTAGTATATTAAACAAGTAAACCTGCTGTATGCTCACCCAGAAAGGAAC	1140
Db	1136	TTGTTTGGTGGCTAGTATATTAAACAAGTAAACCTGCTGTATGCTCACCCAGAAAGGAAC	1195
QY	1141	AGAGCATGTCAGTCTTTGATTCCTGAGAGCCCTTTTTCATTAACAATTCAGAAATAAGTCTGA	1200
Db	1196	AGAGCATGTCAGTCTTTGATTCCTGAGAGCCCTTTTTCATTAACAATTCAGAAATAAGTCTGA	1255

QY	1201	AGGCCACATCATTTTTAGGAAAGAGTTTTAATTCGTAACCTTTAAAGGTCAAAGAGTTATAGA	1260
Db	1256	AGGCCACATCATTTTTAGGAAAGAGTTTTAATTCGTAACCTTTAAAGGTCAAAGAGTTATAGA	1315
QY	1261	TTTCCACTGTATTACCTCGGATACCTCAATTTGAATGTGAGAAATGTCAGAAATGAGGCTATT	1320
Db	1316	TTTCCACTGTATTACCTCGGATACCTCAATTTGAATGTGAGAAATGTCAGAAATGAGGCTATT	1375
QY	1321	ACCTATCAAAAAATCAATGCGCCTCAAGGTGGCACTCTTTTGTACAACTTTATCCGGAGA	1380
Db	1376	ACCTATCAAAAAATCAATGCGCCTCAAGGTGGCACTCTTTTGTACAACTTTATCCGGAGA	1435
QY	1381	CACCTTCTGACACTGAGAGAGGAACCTGTGTGTGCTGAGTTATGGGTTTCTCTTGAATAATG	1440
Db	1436	CACCTTCTGACACTGAGAGAGGAACCTGTGTGTGCTGAGTTATGGGTTTCTCTTGAATAATG	1495
QY	1441	AGCTTTGGACCAAGCTGCTTAACTTAGGAAGGAGCTAATCAGTACCCAGCTAGATCTTGT	1500
Db	1496	AGCTTTGGACCAAGCTGCTTAACTTAGGAAGGAGCTAATCAGTACCCAGCTAGATCTTGT	1555
QY	1501	CCTGTGCCAAAAAGTTATACATCCATCTTTTGAAGCAGTTTCTCAATATGATCGTATTAT	1560
Db	1556	CCTGTGCCAAAAAGTTATACATCCATCTTTTGAAGCAGTTTCTCAATATGATCGTATTAT	1615
QY	1561	TGCCATAGACAGAAATTTGGAGTGACCTCTGATGGAAACCCCTGACTAAAATGACAGGAACACA	1620
Db	1616	TGCCATAGACAGAAATTTGGAGTGACCTCTGATGGAAACCCCTGACTAAAATGACAGGAACACA	1675
QY	1621	GGCTATTGGATCCCTAGGCTCAATATGCTTAAATAGTTATGGAAGTGTGAAAGATGTGTG	1680
Db	1676	GGCTATTGGATCCCTAGGCTCAATATGCTTAAATAGTTATGGAAGTGTGAAAGATGTGTG	1735
QY	1681	CACCTGAAAAATTTGGCTCCCAACATTTTTCATCTTATCTTAATGAAGCAACAATCTG	1740
Db	1736	CACCTGAAAAATTTGGCTCCCAACATTTTTCATCTTATCTTAAATGAAGCAACAATCTG	1795
QY	1741	CAGCTTGTCTCTGCAACAGAAATGACACTGCGCTGGGATGAGTGAAGCTCACGCTGTCA	1800
Db	1796	CAGCTTGTCTCTGCAACAGAAATGACACTGCGCTGGGATGAGTGAAGCTCACGCTGTCA	1855
QY	1801	GACGCACTGCACTGTCTGCACTTAACTCAAGGAAACCATGGGCTTTGTTGGGAGGTGG	1860
Db	1856	GACGCACTGCACTGTCTGCACTTAACTCAAGGAAACCATGGGCTTTGTTGGGAGGTGG	1915
QY	1861	CTGTACTGAAACTCAATTTGGCTGCAATATACAGACAAGACTACAGACCCAGAAAG	1920
Db	1916	CTGTACTGAAACTCAATTTGGCTGCAATATACAGACAAGACTACAGACCCAGAAAG	1975
QY	1921	CATTTCTCAAGATGATGATGACTCAACAGAACTTCAATTTAAATTTGCTGAAGCATTTTG	1980
Db	1976	CATTTCTCAAGATGATGATGACTCAACAGAACTTCAATTTAAATTTGCTGAAGCATTTTG	2035
QY	1981	CAGTGCCCTTAGAATCTGTTTGGCTCTTTTAGAACATGATGAGGTGAAATTTCTCACTGA	2040
Db	2036	CAGTGCCCTTAGAATCTGTTTGGCTCTTTTAGAACATGATGAGGTGAAATTTCTCACTGA	2095
QY	2041	CATGAAGTATGGAACCTTTTGGTCAAGTTCAGGAGATTCCTCCTGTGTTGCTAACTGGCC	2100
Db	2096	CATGAAGTATGGAACCTTTTGGTCAAGTTCAGGAGATTCCTCCTGTGTTGCTAACTGGCC	2155
QY	2101	AGATTTGCTTTTACAGAGTGGCTGTGGATTAACAAATAGCCAGGAAGAACTCAACTGGTC	2160
Db	2156	AGATTTGCTTTTACAGAGTGGCTGTGGATTAACAAATAGCCAGGAAGAACTCAACTGGTC	2215
QY	2161	TTTCTTAAAGAACACACGTCGTCCATTTTGTGCAAAAAGCTGCCCTTCCACATGAAGCTGT	2220
Db	2216	TTTCTTAAAGAACACACGTCGTCCATTTTGTGCAAAAAGCTGCCCTTCCACATGAAGCTGT	2275
QY	2221	GGGCTCAGCAGCAACCTGACCTTGGACTGTTTGAATGCTGAAAGCTTAGTGGCTACAGGT	2280
Db	2276	GGTCTCAGCAGCAACCTGACCTTGGACTGTTTGAATGCTGAAAGCTTAGTGGCTACAGGT	2335
QY	2281	GGCTGTAGAGACAGCCAAATTTTGATTTGGATCTTTTCATATGTTATTGAAGATAAAAACTA	2340

Db 2336 GCGCTGTAGACAGCCCAATTGATTTTGGATCTTTCATATGTTATTGAAGATAAAACTA 2395
QY 2341 AGAAGATAGCATGTTCTGATTACAGAGAGAAACAAATAAACTAGTCTCTTGGCAATTGAGA 2400
Db 2396 AGAAGATAGCATGTTCTGATTACAGAGAGAAACAAATAAACTAGTCTCTTGGCAATTGAAA 2455
QY 2401 AA 2442
Db 2456 AA 2497

RESULT 3
AF221993 2678 bp mRNA linear PRI 02-JUN-2000
LOCUS Homo sapiens putative chaperonin MKKS (MKKS) mRNA, complete cds,
DEFINITION alternatively spliced.

ACCESSION AF221993
VERSION AF221993.1 GI:8163833
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)

REFERENCE
AUTHORS Stone, D.L., Slavotinek, A., Bouffard, G.G., Banerjee-Basu, S.,
Baxevarian, A.D., Barr, M. and Biesecker, L.G.
TITLE Mutation of a gene encoding a putative chaperonin causes
McKusick-Kaufman syndrome

JOURNAL Nat. Genet. 25 (1), 79-82 (2000)
MEDLINE 20264373
PUBMED 10802661
REFERENCE
AUTHORS Biesecker, L.G., Slavotinek, A.M. and Stone, D.L.
TITLE Direct Submission
JOURNAL Submitted (07-JAN-2000) LGDR, NHGRI/NIH, 49 Convent Drive,
Bethesda, MD 20892-4472, USA

FEATURES
source
1..2678
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/map="20p12"
/chromosome="20"
/notes="corresponds to genomic sequence deposited in
GenBank Accession Number AL034430"

gene
1..2678
/genes="MKKS"
/notes="McKusick-Kaufman (hydrometrocoplos-polydactyly)
syndrome"
888..2600
/gene="MKKS"
/note="transcript MKSB; includes alternative exon 1B"

CDS
888..2600
/product="putative chaperonin MKKS"
/protein_id="AAF73873.1"
/db_xref="GI:8163834"
/translation="MSRLAKKPSLCKSEPLTTERVTTLSVLKRVTSYCGSGLK
QLHNGGGVCTTSQSALLSHLLVTHPIKLTASIQNHVSFSDGFTALLCNL
IENVRLGTPPTVIRLNKHLISLCISYLKSETCGCRIIPVDFSTQILLCVRSILTS
KPAKMLTRKTEHVSAIILRAFLLTIPENAEHGIILGKSLIPLKRGVRIDSTVLP
LIMSVQMLRLPIPKKSTALKVAFCTTSLGSDTGTGTVVSVYGSLENVLDOL
LNGROLISDHDVLCQKVIHPSLKQFLNMHRI IADIRGVTLMPLTKMTGTPIG
SLGSIQNSVGSVKDVTAKFGSKHFLIIPNEATICSILLGNRNDTANDELKLCQT
ALHVQLITLKEPALLGGCTETHLAAYIRKHTNDPESILKDDCTQTELQIIAEF
CSALESVSGLEHDGGEILTDMKYIHLVSWQADSPCVANPDLISQCGLYNSQEEL
NWSFLRSTRPFPQPSCLPHEAVGSASNLTLCLTAKLSGLQAVETANLILDLSTVI
EDKN"

ORIGIN

Query Match 97.6%; Score 2412.2; DB 9; Length 2678;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2414; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 TATGAGCCTTCGGAACTTGTGGAGAGACTACAAAGTTTGTGGTTTATGGTCCCTTTAGT 60
Db 260 TATGAGCCTTCGGAACTTGTGGAGAGACTACAAAGTTTGTGGTTTATGGTCCCTTTAGT 319
QY 61 TGGGCTCATACATTTGGGGTGGTACAGAATCAAAAGCAGCCCTGTTTCCAAATACCTAA 120
Db 320 TGGGCTCATACATTTGGGGTGGTACAGAATCAAAAGCAGCCCTGTTTCCAAATACCTAA 379
QY 121 AAACGACGACATTCCTGAGCAAGATAGTCTGGGACTTTCAAAATCTTCAGAGAGCCCAAT 180
Db 380 AAACGACGACATTCCTGAGCAAGATAGTCTGGGACTTTCAAAATCTTCAGAGAGCCCAAT 439
QY 181 CCAGGGGAAGTAGCAGGCTTGCATCTTCAGGTAAAGAGCAGCTTTGAAATCTCGAGCTTC 240
Db 440 CCAGGGGAAGTAGCAGGCTTGCATCTTCAGGTAAAGAGCAGCTTTGAAATCTCGAGCTTC 499
QY 241 ATATCGAAAGAGAGATGAAATATACAGTTGGATTAGAAAGAACTGGCTTCTGTAGCT 300
Db 500 ATATCGAAAGAGAGATGAAATATACAGTTGGATTAGAAAGAACTGGCTTCTGTAGCT 559
QY 301 GGGATATCTTTCATAGGTGTCATCTTGGAAACATATCTTTTGCAGAGGCTCGCAAGCAG 360
Db 560 GGGATATCTTTCATAGGTGTCATCTTGGAAACATATCTTTTGCAGAGGCTCGCAAGCAG 619
QY 361 TCTGTAAAAATTTTCAGTCTCAAAAGCAACAAAGAGATTTGAAGAGTGAAGTAAAAATAAT 420
Db 620 TCTGTAAAAATTTTCAGTCTCAAAAGCAACAAAGAGATTTGAAGAGTGAAGTAAAAATAAT 679
QY 421 ATTTGGAAATATCAATTTGTCTATTAATCATTTCTATGCTGATTTAGTTCATAAACAATTGA 480
Db 680 ATTTGGAAATATCAATTTGTCTATTAATCATTTCTATGCTGATTTAGTTCATAAACAATTGA 739
QY 481 ACTTTTGTATTTTATAGCCACATGCTGCATATTCATATCTTAAATTTCTTAAAGAAATAAT 540
Db 740 ACTTTTGTATTTTATAGCCACATGCTGCATATTCATATCTTAAATTTCTTAAAGAAATAAT 799
QY 541 TTTTAATGTTAAAAACGTGATAATGCAATAAATAGAAAAATGTGGTTTACAAAAATAAAACG 600
Db 800 TTTTAATGTTAAAAACGTGATAATGCAATAAATAGAAAAATGTGGTTTACAAAAATAAAACG 859
QY 601 GTCTTCATAGTTTACCACTGAAAGTAAAGATGTCTGTTTGGAGCTTAAAGAGCCATCAT 660
Db 860 GTCTTCATAGTTTACCACTGAAAGTAAAGATGTCTGTTTGGAGCTTAAAGAGCCATCAT 919
QY 661 GTGTAAAGTGAACCACTGACAACTGAGAGAGTCAAGGACCACTTTCTGCTTGTGAAGAG 720
Db 920 GTGTAAAGTGAACCACTGACAACTGAGAGAGTCAAGGACCACTTTCTGCTTGTGAAGAG 979
QY 721 AATTGTAACATCATGCTATGGGCCCTCAGGTAGGCTGAAAGCAGCTGCACAATGGCTTTGG 780
Db 980 AATTGTAACATCATGCTATGGGCCCTCAGGTAGGCTGAAAGCAGCTGCACAATGGCTTTGG 1039
QY 781 AGGTAAAGTGTACAACTCAAGTCTCAGCTCTGCTCAGTCACTTTTGGTTCACACA 840
Db 1040 AGGTAAAGTGTACAACTCAAGTCTCAGCTCTGCTCAGTCACTTTTGGTTCACACA 1099
QY 841 TCCCATTTTAAAGATCCTGACAGCCCTCAATACAGATCATGTGTCAAGCTTCAGTCAATTG 900
Db 1100 TCCCATTTTAAAGATCCTGACAGCCCTCAATACAGATCATGTGTCAAGCTTCAGTCAATTG 1159
QY 901 TGGCTTATTCACAGCTATTCTTTTGTGCAACCTGATTGAAAAATGTTTCAGAGATTAGGCTT 960
Db 1160 TGGCTTATTCACAGCTATTCTTTTGTGCAACCTGATTGAAAAATGTTTCAGAGATTAGGCTT 1219
QY 961 GACACCCACCACTGTCAATAGATTAAATAAAACATCTTTTGGAGTCTTTTGCATCAGTTATCT 1020
Db 1220 GACACCCACCACTGTCAATAGATTAAATAAAACATCTTTTGGAGTCTTTTGCATCAGTTATCT 1279
QY 1021 CAAGTCTGATACCTGTGGTTGTGCAATCCAGTGGAGCTTTAGTAGTACTCAGATCTCTCT 1080
Db 1280 CAAGTCTGAGACCTGTGGTTGTGCAATCCAGTGGAGCTTTAGTAGTACTCAGATCTCTCT 1339

QY 1081 TTGTTTGGTGGCTAGTATATTAAACAAGTAAACCTGCTGTATGCTCACACAGAAAGAAAC 1140
Db 1340 TTGTTTGGTGGCTAGTATATTAAACAAGTAAACCTGCTGTATGCTCACACAGAAAGAAAC 1399
QY 1141 AGAGCATGTGTCAGTCTGATCTGAGAGCCTTTTGGCTTACAAATCCAGAAATGCTGA 1200
Db 1400 AGAGCATGTGTCAGTCTGAGAGCCTTTTGGCTTACAAATCCAGAAATGCTGA 1459
QY 1201 AGGCCATCATCTTTTAGGAAAGAGTTTAAATGTCCTTTTAAAGGTCAAAGAGTTATAGA 1260
Db 1460 AGGCCATCATCTTTTAGGAAAGAGTTTAAATGTCCTTTTAAAGGTCAAAGAGTTATAGA 1519
QY 1261 TTCCACTGTATTACCTGGGATCTCAATTGAAATGTCAGAGTTCAATTAATGAGGCTATT 1320
Db 1520 TTCCACTGTATTACCTGGGATCTCAATTGAAATGTCAGAGTTCAATTAATGAGGCTATT 1579
QY 1321 ACCATTCAAAAAATCAACTGCCCTCAAGGTGGGCACCTTTTGTACAACTTTATCCGGAGA 1380
Db 1580 ACCATTCAAAAAATCAACTGCCCTCAAGGTGGGCACCTTTTGTACAACTTTATCCGGAGA 1639
QY 1381 CACTTCTGACATGGAAGAAAGAACTGTGTGGTCACTTATGAGGTTTCTCTTGAATAATGC 1440
Db 1640 CACTTCTGACATGGAAGAAAGAACTGTGTGGTCACTTATGAGGTTTCTCTTGAATAATGC 1699
QY 1441 AGTCTTGGACCAAGCTGCTTAACTAGGAAGGCAGCTAAATCAGTGACCAAGTATGCTGT 1500
Db 1700 AGTCTTGGACCAAGCTGCTTAACTAGGAAGGCAGCTAAATCAGTGACCAAGTATGCTGT 1759
QY 1501 CCTGTGCCAAAAAGTTATACATCCATCTTTTGAAGCAGTTTCTCAATATCATCGTATTAT 1560
Db 1760 CCTGTGCCAAAAAGTTATACATCCATCTTTTGAAGCAGTTTCTCAATATCATCGTATTAT 1819
QY 1561 TGCCATPAGACAGAAATGGAGTGACTCTGTATGGAACCCCTGACTTAAATGACAGGAACACA 1620
Db 1820 TGCCATPAGACAGAAATGGAGTGACTCTGTATGGAACCCCTGACTTAAATGACAGGAACACA 1879
QY 1621 GCCTATTGATCCCTAGGCTCAATATGCTCTAATAGTTATGGAAGTGAAGATGTGTG 1680
Db 1880 GCCTATTGATCCCTAGGCTCAATATGCTCTAATAGTTATGGAAGTGAAGATGTGTG 1939
QY 1681 CACTGCAAAATTTGGCTCCCAACATTTTTCATCTTATCTTCTAATGAAGCAACAATCTG 1740
Db 1940 CACTGCAAAATTTGGCTCCCAACATTTTTCATCTTATCTTCTAATGAAGCAACAATCTG 1999
QY 1741 CAGCTTGCTTCTCTGCAACAGAAATGACACTGCTGGGATGAGCTGAAGCTCACGTGTCA 1800
Db 2000 CAGCTTGCTTCTCTGCAACAGAAATGACACTGCTGGGATGAGCTGAAGCTCACGTGTCA 2059
QY 1801 GACGGCACTGCATGCTCTGCAGTTTACACTCAAGGAACCATGGCTTTCTGGGAGGTGG 1860
Db 2060 GACGGCACTGCATGCTCTGCAGTTTACACTCAAGGAACCATGGCTTTCTGGGAGGTGG 2119
QY 1861 CTGTACTGAAACTCATTTGGCTGCATATATCAGACACAGACTCACAACGACCCAGAAAG 1920
Db 2120 CTGTACTGAAACTCATTTGGCTGCATATATCAGACACAGACTCACAACGACCCAGAAAG 2179
QY 1921 CATTTCTCAAGATGATGAATGACTCAACAGAACTTCAATTAATTTGCTGAAGCATTTTG 1980
Db 2180 CATTTCTCAAGATGATGAATGACTCAACAGAACTTCAATTAATTTGCTGAAGCATTTTG 2239
QY 1981 CAGTGCCCTAGATCTGTTTGGCTCTTTAGAACATGATGGAGGTGAATTTCTCACTCA 2040
Db 2240 CAGTGCCCTAGATCTGTTTGGCTCTTTAGAACATGATGGAGGTGAATTTCTCACTCA 2099
QY 2041 CATGAAGTATGGAACACTTTGGTCAGTTTCAAGGAGATTTCTCCTGTGTTGCTAACTGGCC 2100
Db 2300 CATGAAGTATGGAACACTTTGGTCAGTTTCAAGGAGATTTCTCCTGTGTTGCTAACTGGCC 2359
QY 2101 AGATTTGCTTTTCAAGTGGCTGTGGATATATCAATAGCAGGAAGAACTCAACTGCTGTC 2160
Db 2360 AGATTTGCTTTTCAAGTGGCTGTGGATATATCAATAGCAGGAAGAACTCAACTGCTGTC 2419
QY 2161 TTTTCTTAAGAGCACAGCTGCTCCATTTTGTGCCACAAAGCTGCTTCCACATGAAGCTGT 2220

Db 2420 TTTCTTAAGAACACACAGCTGCTCCATTTTGTGCCACAAAGCTGCCCTTCCACATGAAGCTGT 2479
QY 2221 GGGCTCAGCCAGCAACCTGACCTTGGACTGTTTGAAGCTGCTCAAGAGCTTAGTGGCTACAGGT 2280
Db 2480 GGGCTCAGCCAGCAACCTGACCTTGGACTGTTTGAAGCTGCTCAAGAGCTTAGTGGCTACAGGT 2539
QY 2281 GGGCTCAGCCAGCAACCTGACCTTGGACTGTTTGAAGCTGCTCAAGAGCTTAGTGGCTACAGGT 2340
Db 2540 GGGCTCAGCCAGCAACCTGACCTTGGACTGTTTGAAGCTGCTCAAGAGCTTAGTGGCTACAGGT 2599
QY 2341 AGAGAAATGACATGCTTGGTATTACAGAGAAACAAATAAAGTCTGTTGGCAATTTGAGA 2400
Db 2600 AGAGAAATGACATGCTTGGTATTACAGAGAAACAAATAAAGTCTGTTGGCAATTTGAGA 2659
QY 2401 AAAAAAAAAAAAAAAAAAAAAA 2419
Db 2660 AAAAAAAAAAAAAAAAAAAAAA 2678

RESULT 4
HSM802462
LOCUS HSM802462 2512 bp mRNA linear PRI 18-FEB-2000
DEFINITION Homo sapiens mRNA; cDNA DKFZp761A072 (from clone DKFZp761A072).
ACCESSION AL157427
VERSION AL157427.1 GI:7018456
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2512)
AUTHORS Ottenwaelder,B., Obermaier,B., Mewes,H.W., Weil,B. and Wiemann,S.
TITLE Direct Submission
JOURNAL Submitted (15-FEB-2000) MIPS, Am Klopferspitz 18a, D-82152
Martinsried, GERMANY
COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; sequenced by MedGenomix (Martinsried/Germany) within the cDNA sequencing consortium of the German Genome Project. This clone (DKFZp761A072) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further information about the clone and the sequencing project is available at http://www.mips.biochem.mpg.de/proj/cDNA/.

FEATURES
Location/Qualifiers
1..2512
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/map="32.60 cR from top of Chr20 linkage group"
/clone="DKFZp761A072"
/tissue_type="amygdala"
/clone_lib="761 (synonym: hamy2). Vector pSport1; host DH10B; sites NotI + SalI"
/dev_stage="adult"
1..2512
/gene="DKFZp761A072"
721..2433
/gene="DKFZp761A072"
/note="unknown protein"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAB75652.1"
/db_xref="GI:7018457"
/db_xref="GOA:O9NPJ1"
/db_xref="UniProt/Swiss-Prot:O9NPJ1"
/translation="MSRLAEAKPSLCKSEPLTTERVRTLVLKIVTSCYSPSRLK QLNHGGYVCTTSQSALLSHLVTHPIKLITASIQNHVSSPDSGGLFTALCCNL IENVQLGLTPTTVIRLNKHLISLCISYKSETCGRIPVPFSSQIILLCLVRSILTS KPACMLTKRETHVSALIRAPLITIPENABGHIILGKSLIVPLQQRVIDSTVLPGL LTEMSEVOLMRLLPIKSKTALKVALFCTTISGDTSDTGTGEGTVVWSYGVSLNVALDQL LNLGRQLISDHVDLVLCQVIHPSLKQFLNMRHRIIADIRIGVTLMEPLTKMTGTQPIG

SLGSI CPNSYGSVKDVCTAKFPGSKHPLHLPNEATICSLLLCNRNPTAWDELKLTQOT
ALHVLQTLKEPWLALGGCTETHLAAYIRKHNDPESILKDDKCTQTELOLIARAF
CSALESVVGSLHDGGBILDMKYGHILMSVQADSPCVANWPDLLSCGCGLYNSOEL
NWSFLRSTRPFPQSCLPHEAVGSANLTLDCILAKLSGLQVAVETANLILDLISYVI
EDKN"
polyA_signal 2466..2471
/gene="DKFZp761A072"
polyA_site 2489
/gene="DKFZp761A072"
ORIGIN

Query Match 97.6%; Score 2412; DB 9; Length 2512;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2415; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1 TATGAGCCTTCGGAACCTTGTGGAGAGACTACAAAGTTTTCGTTGTTATGCTCCCTTTAGT 60
DB 93 TATGAGCCTTCGGAACCTTGTGGAGAGACTACAAAGTTTTCGTTGTTATGCTCCCTTTAGT 152
QY 61 TGGGCTCATACATTTGGGGTGGTACAGAATCAAAAGCAGCCCTGTTTTCAAAATACCTAA 120
DB 153 TGGGCTCATACATTTGGGGTGGTACAGAATCAAAAGCAGCCCTGTTTTCAAAATACCTAA 212
QY 121 AAACGACGACATTCCTGAGCAAGATAGTCTGGGACCTTTCAAATCTTCAGAAGAGCCAAAT 180
DB 213 AAACGACGACATTCCTGAGCAAGATAGTCTGGGACCTTTCAAATCTTCAGAAGAGCCAAAT 272
QY 181 CCAGGGGAAGTACGAGCTTGCATCTTCAGGTAAAGAACGACCTTTGAATCTGAGCTTC 240
DB 273 CCAGGGGAAGTACGAGCTTGCATCTTCAGGTAAAGAACGACCTTTGAATCTGAGCTTC 332
QY 241 ATATCGAAAGAGAGATGAAATATACAGTTGGATTAGAAAGAACTGGCTTCCTGTAGCT 300
DB 333 ATATCGAAAGAGAGATGAAATATACAGTTGGATTAGAAAGAACTGGCTTCCTGTAGCT 392
QY 301 GGGATATCTTTCTATAGGTGTCATCTTGGAAACATACATCTTTTGGAGAGGTCTGCAAGCAG 360
DB 393 GGGATATCTTTCTATAGGTGTCATCTTGGAAACATACATCTTTTGGAGAGGTCTGCAAGCAG 452
QY 361 TCTGTAATAATTTTCAGTCTCAAGCAACAAAGAGATTTGAAGAGTGAAGTAAATTAAT 420
DB 453 TCTGTAATAATTTTCAGTCTCAAGCAACAAAGAGATTTGAAGAGTGAAGTAAATTAAT 512
QY 421 ATTTGGAAATPACTAATTTGTCAATTAATCATCTATGCTGATTAAGTCTTCAATAACATTTGA 480
DB 513 ATTTGGAAATPACTAATTTGTCAATTAATCATCTATGCTGATTAAGTCTTCAATAACATTTGA 572
QY 481 ACTTTTGTATTTATAGCCACAATGCTGCAATTTCAATTTAATTTCTTAAGAAATAATTT 540
DB 573 ACTTTTGTATTTATAGCCACAATGCTGCAATTTCAATTTAATTTCTTAAGAAATAATTT 632
QY 541 TTTTAATGTTAAACGTTGATATGCAATTAATAGAAAATGTTGTTTACAAAATAAAACG 600
DB 633 TTTTAATGTTAAACGTTGATTAATGCAATTAATAGAAAATGTTGTTTACAAAATAAAACG 692
QY 601 GTCTTCACTAGTTACACCTGAAAGTAAAGTGTCTCGTTTGGAAAGCTAAGAGCCATCATTT 660
DB 693 GTCTTCACTAGTTACACCTGAAAGTAAAGTGTCTCGTTTGGAAAGCTAAGAGCCATCATTT 752
QY 661 GTCTAAGAGTGAACCACTGACAACTGAGAGAGTCAAGACCACTTTCTGTCTTGGAAAAG 720
DB 753 GTCTAAGAGTGAACCACTGACAACTGAGAGAGTCAAGACCACTTTCTGTCTTGGAAAAG 812
QY 721 AATTGTAACATCATGCTATGCCCCCTCAGTAGAGCTGAAGCAGCTGACAAATGGCTTTGG 780
DB 813 AATTGTAACATCATGCTATGCCCCCTCAGTAGAGCTGAAGCAGCTGACAAATGGCTTTGG 872
QY 781 AGGTTAGCTGTGTACAACTCACAAGTCTCAGCTCTGCTCAGTCACCTTTTGTGTGTCACACA 840
DB 873 AGGTTAGCTGTGTGTACAACTCACAAGTCTCAGCTCTGCTCAGTCACCTTTTGTGTGTCACACA 932
QY 841 TCCCATTTTAAAGATCCTGACAGCCTCCATACAGAATCATGTGTCAGGCTTCAGTGATTTG 900

DB 933 TCCCATTTTAAAGATCCTGACAGCCTCCATACAGAATCATGTGTCAGGCTTCAGTGATTTG 992
QY 901 TGGCTTATTCACAGCTATTTCTTTTGTGCGAACCTGATTGAAAATGTTTCAGAGATTAGGCTT 960
DB 993 TGGCTTATTCACAGCTATTTCTTTTGTGCGAACCTGATTGAAAATGTTTCAGAGATTAGGCTT 1052
QY 961 GACACCCACCACTGTCAITTAGATTAAATAAACATCTTTTGTGAGTCTTTTGCATCAGTTATCT 1020
DB 1053 GACACCCACCACTGTCAITTAGATTAAATAAACATCTTTTGTGAGTCTTTTGCATCAGTTATCT 1112
QY 1021 CAAGTCTGATACCTGTGGTTGTGCAATCCAGTGGAGCTTTTAGTAGTACTCAGATCCCTCT 1080
DB 1113 CAAGTCTGAGACCTGTGGTTGTGCAATCCAGTGGAGCTTTTAGTAGTACTCAGATCCCTCT 1172
QY 1081 TTGCTTTGTGCGTAGTATATTAAACAAGTAAACCTGCTGTATGCTCACCAGAAAGGAAC 1140
DB 1173 TTGCTTTGTGCGTAGTATATTAAACAAGTAAACCTGCTGTATGCTCACCAGAAAGGAAC 1232
QY 1141 AGAGCATGTCAAGTCTTTGATCCTGAGAGCCTTTTGTCTTACAAATTCAGAAAATGCTGA 1200
DB 1233 AGAGCATGTCAAGTCTTTGATCCTGAGAGCCTTTTGTCTTACAAATTCAGAAAATGCTGA 1292
QY 1201 AGGCCACATCATTTTAGGAAAAGATTTAAATTTGATCCTTTAAAGGTCAAGAGTTATAGA 1260
DB 1293 AGGCCACATCATTTTAGGAAAAGATTTAAATTTGATCCTTTAAAGGTCAAGAGTTATAGA 1352
QY 1261 TTCCACTGTATTACCTGGGATCTCAATGAAAATGTCAGAGTTCAATTAATGAGGCTATT 1320
DB 1353 TTCCACTGTATTACCTGGGATCTCAATGAAAATGTCAGAGTTCAATTAATGAGGCTATT 1412
QY 1321 ACCTATCAAAAAATCAACTGCCCTCAAGGTGGGACCTCTTTTGTACAACTTTATCCGGAGA 1380
DB 1413 ACCTATCAAAAAATCAACTGCCCTCAAGGTGGGACCTCTTTTGTACAACTTTATCCGGAGA 1472
QY 1381 CACTTCTGACACTGGAGAAAGGAACTGTGTGTGTCAGTTATGGGTTTCTCTTCAAAATGC 1440
DB 1473 CACTTCTGACACTGGAGAAAGGAACTGTGTGTGTCAGTTATGGGTTTCTCTTCAAAATGC 1532
QY 1441 AGTCTTGGACCACTGCTTAACTAGGAAGGAGCTTAATCAGTGACCACTAGATCTTTGT 1500
DB 1533 AGTCTTGGACCACTGCTTAACTAGGAAGGAGCTTAATCAGTGACCACTAGATCTTTGT 1592
QY 1501 CCTGTGCCAAAAAGTTATACATCCATCTTTTGAAGCAGTTTCTCAATATGATCGTATTAT 1560
DB 1593 CCTGTGCCAAAAAGTTATACATCCATCTTTTGAAGCAGTTTCTCAATATGATCGTATTAT 1652
QY 1561 TGGCATAGACAGAAATTTGGAGTGACTCTGTATGGAAACCCCTGACTAAAATGACAGAAACACA 1620
DB 1653 TGGCATAGACAGAAATTTGGAGTGACTCTGTATGGAAACCCCTGACTAAAATGACAGAAACACA 1712
QY 1621 GCCTATTGGATCCCTAGGCTCAATATGCTCCTAATAGTTATGGAGTGTGAAAGATGTGTG 1680
DB 1713 GCCTATTGGATCCCTAGGCTCAATATGCTCCTAATAGTTATGGAGTGTGAAAGATGTGTG 1772
QY 1681 CACTGCAAAATTTGGCTCCCAACATTTTCTTATCTTATTTCTTAATGAAAGCAACATCTG 1740
DB 1773 CACTGCAAAATTTGGCTCCCAACATTTTCTTATCTTATTTCTTAATGAAAGCAACATCTG 1832
QY 1741 CAGCTTGTCTTCTGCAACAGAAAATGACACTGCTGGGATGAGCTGAAGCTCACGCTGTCA 1800
DB 1833 CAGCTTGTCTTCTGCAACAGAAAATGACACTGCTGGGATGAGCTGAAGCTCACGCTGTCA 1892
QY 1801 GAGGCACTGCAATGCTCTGCACTTAACACTCAAGGAAACCATGGGCTTTGTTGGAGGTGG 1860
DB 1893 GAGGCACTGCAATGCTCTGCACTTAACACTCAAGGAAACCATGGGCTTTGTTGGAGGTGG 1952
QY 1861 CTGTACTGAAACTCATTTTGGCTGCAATATCAGACAAAGACTCACAAACGACCCAGAAAG 1920
DB 1953 CTGTACTGAAACTCATTTTGGCTGCAATATCAGACAAAGACTCACAAACGACCCAGAAAG 2012
QY 1921 CATTTCTCAAGAGATGAATGTACTCAAAACAGAACTTCAATTAATTTGCTGAAGCACTTTTG 1980
DB 2013 CATTTCTCAAGAGATGAATGTACTCAAAACAGAACTTCAATTAATTTGCTGAAGCACTTTTG 2072

QY 1981 CAGTGCCTAGATCTGTGTGGCTCTTTAGAACATGATGAGGTGAAATTCCTACTGA 2040
Db 2073 CAGTGCCTAGATCTGTGTGGCTCTTTAGAACATGATGAGGTGAAATTCCTACTGA 2132
QY 2041 CATGAAGTATGGACACCTTTGGTCAGTTCAGGCAGATTCCTCCTGTGTGCTAACTGGCC 2100
Db 2133 CATGAAGTATGGACACCTTTGGTCAGTTCAGGCAGATTCCTCCTGTGTGCTAACTGGCC 2192
QY 2101 AGATTTGCTTTACAGTGGCTGTGGATATATCAATACCCAGGAAGAACTCAACTGGTC 2160
Db 2193 AGATTTGCTTTACAGTGGCTGTGGATATATCAATACCCAGGAAGAACTCAACTGGTC 2252
QY 2161 TTTCTTAAGACAGACGCTGCTCCATTTGTGCCAAGAGCTGCTTCCACATGAAGCTGT 2220
Db 2253 TTTCTTAAGACAGACGCTGCTCCATTTGTGCCAAGAGCTGCTTCCACATGAAGCTGT 2312
QY 2221 GGCTCAGCCAGCAACCTGACCTTGGACTGTTTGACTGCAGAACTTAGTGGCTACAGGT 2280
Db 2313 GGCTCAGCCAGCAACCTGACCTTGGACTGTTTGACTGCAGAACTTAGTGGCTACAGGT 2372
QY 2281 GGCTGTAGACAGCCAAATTTGATTTGGGATCTTTCATATGTTTAAAGATAAAACTA 2340
Db 2373 GGCTGTAGACAGCCAAATTTGATTTGGGATCTTTCATATGTTTAAAGATAAAACTA 2432
QY 2341 AGNATAGCATGTTTCGTATTACAGAGAAACAAATAAACTAGTCTGTTGGCNAATTGAGA 2400
Db 2433 AGAATAGCATGTTTCGTATTACAGAGAAACAAATAAACTAGTCTGTTGGCNAATTGAAA 2492
QY 2401 AAAAAAAAAAAAAAAAAAAAA 2420
Db 2493 AAAAAAAAAAAAAAAAAAAAA 2512

RESULT 5
AF221992 2516 bp mRNA linear PRI 02-JUN-2000
LOCUS Homo sapiens putative chaperonin MKKS (MKKS) mRNA, complete cds,
DEFINITION alternatively spliced.
ACCESSION AF221992
VERSION AF221992.1 GI:8163831
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2516)
AUTHORS Stone,D.L., Slavotinek,A., Bouffard,G.G., Banerjee-Basu,S.,
Baxevarian,A.D., Barr,M. and Biesecker,L.G.
TITLE Mutation of a gene encoding a putative chaperonin causes
McKusick-Kaufman syndrome
JOURNAL Nat. Genet. 25 (1), 79-82 (2000)
MEDLINE 20264373
PUBMED 10802661
REFERENCE 2 (bases 1 to 2516)
AUTHORS Biesecker,L.G., Slavotinek,A.M. and Stone,D.L.
TITLE Direct submission
JOURNAL Submitted (07-JAN-2000) LGDR, NHGRI/NIH, 49 Convent Drive,
Bethesda, MD 20892-4472, USA
FEATURES
source
1. .2516
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="20"
/map="20p12"
/note="corresponds to genomic sequence deposited in
GenBank Accession Number AL034430"
1. .2516
/gene="MKKS"
/notes="McKusick-Kaufman (hydrometrocophlos-polydactyly)
syndrome"
736. .2448

/gene="MKKS"
/note="transcript MKKSa; includes alternative exon 1A"
/codon_start=1
/product="putative chaperonin MKKS"
/protein_id="AAF73872.1"
/db_xref="GI:8163832"
/translation="MSRLEAKKPSLCKSEPLTTERVRTLSVLKRIVTSYGPSRLK
QLHNGFGYVCTTSOSSALLSHLLVTHPIKLITASIQNHVSPFSDGGLFTAILCCNL
IENVORLGLTPTTVIRLNKHLUSLCISYKSETCGCRIPVDFSSFOILLCILSVILTS
KPCAMTRKRETHEVSALILRAFLIIPENAEHIIILKSLIVPLKQGVDSVILPGL
LIEMSEVOLMRLLPIPKSTALKVALFCTTSGDSDTGTGTVVGVSVLENVADQL
NLNGRLISDHVDLVLCOKVHPSLKQPLNMHRIIDRIGVTLMPLTKMTGTQPIG
SIGSTCPNSYGVKDVCTAKFGSKHFLIPNEATISILLCNRDTDWDELKTCOT
ALHVLQILKEPALLGGCTETHLAAYIRKHTNDPESILKDDSCCTQELQIAEAF
CSALRSVVGSEHDEGELTDNKGHLWSVQADSPCVANWPDLLSOCGGLNSQEEL
NWSFLRSTRPFPVQSLPHEAVGSASNLTLDCLTAKLSGLQAVAVETANLILDLSTVI
EDKN"

ORIGIN
Query Match 97.1%; Score 2401; DB 9; Length 2516;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2404; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 TATGAGCCTTTCGGAACCTTGTGAGAGACTACAAAGTTTTTGGTTGTTATGGTCCCTTAGT 60
Db 108 TATGAGCCTTTCGGAACCTTGTGAGAGACTACAAAGTTTTTGGTTGTTATGGTCCCTTAGT 167
QY 61 TGGGCTCATACATTTGGGGTGGTACAGAAATCAAAAGCAGCCCTGTTTCCAAATACCTAA 120
Db 168 TGGGCTCATACATTTGGGGTGGTACAGAAATCAAAAGCAGCCCTGTTTCCAAATACCTAA 227
QY 121 AAACGACGACATTCCTGAGCAAGATAGCTGGGACTTTCAAATCTTCAGAAAGCCAAAT 180
Db 228 AAACGACGACATTCCTGAGCAAGATAGCTGGGACTTTCAAATCTTCAGAAAGCCAAAT 287
QY 181 CCAGGGAGTAGCAGGCTTGCAATCTTCAGGTAAAGAGCAGCTTTTGAATCTTGAGCTTC 240
Db 288 CCAGGGAGTAGCAGGCTTGCAATCTTCAGGTAAAGAGCAGCTTTTGAATCTTGAGCTTC 347
QY 241 ATATCGAAAGAGAGATGAAAATACCAAGTTGGAATAGAAAGAACTGGCTTTCTGAGCT 300
Db 348 ATATCGAAAGAGAGATGAAAATACCAAGTTGGAATAGAAAGAACTGGCTTTCTGAGCT 407
QY 301 GGGATATCTTTCATAGGTGTCATCTTGGAACTACTTTTTCGAGAGGCTCCGAAAGCAG 360
Db 408 GGGATATCTTTCATAGGTGTCATCTTGGAACTACTTTTTCGAGAGGCTCCGAAAGCAG 467
QY 361 TCTGTAAAATTTTCAGTCTCAAGCAACAAAGAGATTTGAAGAGTGAAAGTAAAAATAAT 420
Db 468 TCTGTAAAATTTTCAGTCTCAAGCAACAAAGAGATTTGAAGAGTGAAAGTAAAAATAAT 527
QY 421 ATTTGGAAATTAATAATTTGTCATTAATCATTTCTATGCTGATAGCTTCATAAACATTGA 480
Db 528 ATTTGGAAATTAATAATTTGTCATTAATCATTTCTATGCTGATAGCTTCATAAACATTGA 587
QY 481 ACTTTTGTATTTTATAGCCAAATGCTGCATATTCATATCTTAATCTCTAAAGATAAAT 540
Db 588 ACTTTTGTATTTTATAGCCAAATGCTGCATATTCATATCTTAATCTCTAAAGATAAAT 647
QY 541 TTTTAATGTTAAAAACGTGATAATGCAATAAATAGAAAAATGTGTTTTACAAAAATAAAACG 600
Db 648 TTTTAATGTTAAAAACGTGATAATGCAATAAATAGAAAAATGTGTTTTACAAAAATAAAACG 707
QY 601 GTCTTCACTAGTTACCACTGAAAGTAAGATGTCTCGTTTGGAGCTGAAGAGCCATCAT 660
Db 708 GTCTTCACTAGTTACCACTGAAAGTAAGATGTCTCGTTTGGAGCTGAAGAGCCATCAT 767
QY 661 GTGTAAGAGTGAACCACTGACAACTGAGAGAGCTCAGGACCACTTCTGCTCTTGAAGAG 720
Db 768 GTGTAAGAGTGAACCACTGACAACTGAGAGAGCTCAGGACCACTTCTGCTCTTGAAGAG 827
QY 721 AATTGTAAACATCATGCTATGGCCCTTCAGGTAGGCTGAAAGAGCTGCACAAATGGCTTTGG 780

Db 828 AATTGTAACATCATGCTATGGCCCCCTCAGGTAGGCTGAAGCAGCTGCACAATGGCTTTGG 887
Qy 781 AGGTTACGTTGTGTAAACCTCA CAGTCTCTCAGCTCTGCTCAGTCAACCTTTTGGTCAACA 840
Db 888 AGGTTACGTTGTGTAAACCTCA CAGTCTCTCAGCTCTGCTCAGTCAACCTTTTGGTCAACA 947
Qy 841 TCCCAATTTAAAGATTCCTGACAGCTCCATACAGAAATCATGTGTCAAGCTTCAGTGATTG 900
Db 948 TCCCAATTTAAAGATTCCTGACAGCTCCATACAGAAATCATGTGTCAAGCTTCAGTGATTG 1007
Qy 901 TGGCTTATTCACAGCTATCTTTCTGCTCAACCTGATTTGAAATGTTCCAGAGATTAGGCTT 960
Db 1008 TGGCTTATTCACAGCTATCTTTCTGCTCAACCTGATTTGAAATGTTCCAGAGATTAGGCTT 1067
Qy 961 GACACCCACCACTGTCTATTAGATTAATAAATCAATCTTTTGGATCTTTTGGCATCAGTTATCT 1020
Db 1068 GACACCCACCACTGTCTATTAGATTAATAAATCAATCTTTTGGATCTTTTGGCATCAGTTATCT 1127
Qy 1021 CAAGTCTGATACCTGTGGTTGTGCAATCCCAAGTGGACTTTTAGTAGTACTCAGATCTCTCT 1080
Db 1128 CAAGTCTGAGACCTGTGGTTGTGCAATCCCAAGTGGACTTTTAGTAGTACTCAGATCTCTCT 1187
Qy 1081 TTGTTTGGTGTAGTATATTAACAAGTAAACCTGCTGTATGCTCAACCAAGGAAGAAC 1140
Db 1188 TTGTTTGGTGTAGTATATTAACAAGTAAACCTGCTGTATGCTCAACCAAGGAAGAAC 1247
Qy 1141 AGAGCATGTCTAGTCTTTTGTATCTCTGAGAGCTTTTGTCTTACAATCCAGAAATGCTGA 1200
Db 1248 AGAGCATGTCTAGTCTTTTGTATCTCTGAGAGCTTTTGTCTTACAATCCAGAAATGCTGA 1307
Qy 1201 AGGCCATCATTTTAGGAAAGAGTTAAATGTACCTTTTAAAGGTCAAAGAGTTATAGA 1260
Db 1308 AGGCCATCATTTTAGGAAAGAGTTAAATGTACCTTTTAAAGGTCAAAGAGTTATAGA 1367
Qy 1261 TTCACCTGATATACCTGGGTACTCATTTGAAATGTCAGAGTTCAATTAATAGGCTATT 1320
Db 1368 TTCACCTGATATACCTGGGTACTCATTTGAAATGTCAGAGTTCAATTAATAGGCTATT 1427
Qy 1321 ACCTATCAAAAAATCAACTGCCCTCAAGGTGGCACTCTTTTGTACAACTTTATCCGGAGA 1380
Db 1428 ACCTATCAAAAAATCAACTGCCCTCAAGGTGGCACTCTTTTGTACAACTTTATCCGGAGA 1487
Qy 1381 CACTTCTGACATGGAGAGAACTGTGGTGGTCAAGTTATGGGGTTCTCTTGAANAATGC 1440
Db 1488 CACTTCTGACATGGAGAGAACTGTGGTGGTCAAGTTATGGGGTTCTCTTGAANAATGC 1547
Qy 1441 AGTCTTGGACAGCTGCTTAACCTAGGAAGGCACTTAATCAGTGACCACTAGATCTTGT 1500
Db 1548 AGTCTTGGACAGCTGCTTAACCTAGGAAGGCACTTAATCAGTGACCACTAGATCTTGT 1607
Qy 1501 CCTGTGCCAAAAAGTTATACATCATCTTTTGAAGCAGTTTCTCAATATGCATCGTATTAT 1560
Db 1608 CCTGTGCCAAAAAGTTATACATCATCTTTTGAAGCAGTTTCTCAATATGCATCGTATTAT 1667
Qy 1561 TGGCATAGACAGAAATGGAGTGACTCTGATGGAAACCCCTGACTTAAATAGCAGGAACACA 1620
Db 1668 TGGCATAGACAGAAATGGAGTGACTCTGATGGAAACCCCTGACTTAAATAGCAGGAACACA 1727
Qy 1621 GCCTATTGGATCCCTAGGCTCAATATGCTCTTAATAGTTATGGAAGTGAAGATGTGTG 1680
Db 1728 GCCTATTGGATCCCTAGGCTCAATATGCTCTTAATAGTTATGGAAGTGAAGATGTGTG 1787
Qy 1681 CACTGCAAAATTTGGCTCCCAACATTTTTTTTTCATCTTATTCTTAATGAAGCAACAATCTG 1740
Db 1788 CACTGCAAAATTTGGCTCCCAACATTTTTTTTTCATCTTATTCTTAATGAAGCAACAATCTG 1847
Qy 1741 CAGCTTGCTCTCTGCAACAGAAATGACACTGCTGGATGAGCTGAAGCTCAGCTGTCA 1800
Db 1848 CAGCTTGCTCTCTGCAACAGAAATGACACTGCTGGATGAGCTGAAGCTCAGCTGTCA 1907
Qy 1801 GAGGGCACTGCATGCTCTGAGTTAACTCAAGGAACCAATGGCTTTTGGAGAGGTGG 1860
Db 1908 GAGGGCACTGCATGCTCTGAGTTAACTCAAGGAACCAATGGCTTTTGGAGAGGTGG 1967

Qy 1861 CTGTACTGAAACTCATTTGGCTGCATATATCAGACACAGACTCACAACGACCCAGAAAG 1920
Db 1968 CTGTACTGAAACTCATTTGGCTGCATATATCAGACACAGACTCACAACGACCCAGAAAG 2027
Qy 1921 CATTTCTCAAAAGATGATGAATGTACTCAACAGAACTTCAATTAATTTGCTGAAGCATTTTG 1980
Db 2028 CATTTCTCAAAAGATGATGAATGTACTCAACAGAACTTCAATTAATTTGCTGAAGCATTTTG 2087
Qy 1981 CAGTGGCCCTAGAAATCTTTTGGCTCTTTAGAACATGATGGAGGTGAANAATTTCTCACTGA 2040
Db 2088 CAGTGGCCCTAGAAATCTTTTGGCTCTTTAGAACATGATGGAGGTGAANAATTTCTCACTGA 2147
Qy 2041 CATGAAGTATGGACACCTTTTGGTCACTTCAAGGCAGATTTCTCCCTGTGTGTCTAACTGGCC 2100
Db 2148 CATGAAGTATGGACACCTTTTGGTCACTTCAAGGCAGATTTCTCCCTGTGTGTCTAACTGGCC 2207
Qy 2101 AGATTTGCTTTTACAGTGTGGCTGTGGATTTATACAATAGCAGGAAGAACTCAACTGGTC 2160
Db 2208 AGATTTGCTTTTACAGTGTGGCTGTGGATTTATACAATAGCAGGAAGAACTCAACTGGTC 2267
Qy 2161 TTTCCTTAAAGAACACACGCTGCTCCATTTTGTGGCCACAAAGCTGCTTCCCATGAAAGTGT 2220
Db 2268 TTTCCTTAAAGAACACACGCTGCTCCATTTTGTGGCCACAAAGCTGCTTCCCATGAAAGTGT 2327
Qy 2221 GGGCTCAGCCAGCAACCTGACCTTGGACTTGTGACTGCAAAAGCTTAGTGGCTTACAGGT 2280
Db 2328 GGGCTCAGCCAGCAACCTGACCTTGGACTTGTGACTGCAAAAGCTTAGTGGCTTACAGGT 2387
Qy 2281 GGGCTCAGCCAGCAACCTGACCTTGGACTTGTGACTTGTGACTGCAAAAGCTTAGTGGCTTACAGGT 2340
Db 2388 GGGCTCAGCCAGCAACCTGACCTTGGACTTGTGACTTGTGACTGCAAAAGCTTAGTGGCTTACAGGT 2447
Qy 2341 AGAGATAGCATGTTTGGTATTACAGAGAAACAAATAAACTAGTCTGTTGGCAATTGAGA 2400
Db 2448 AGAGATAGCATGTTTGGTATTACAGAGAAACAAATAAACTAGTCTGTTGGCAATTGAGA 2507
Qy 2401 AAAAAAAAAA 2409
Db 2508 AAAAAAAAAA 2516

RESULT 6

AK025741
LOCUS Homo sapiens cDNA: FLJ22088 fis, clone HEP15963.
DEFINITION Homo sapiens cDNA: FLJ22088 fis, clone HEP15963.
ACCESSION AK025741
VERSION AK025741.1 GI:10438352
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
REFERENCE
AUTHORS Kawabata, A., Hikiji, T., Kobatake, N., Inagaki, H., Ikema, Y., Okamoto, S., Okitani, R., Ota, T., Suzuki, Y., Obayashi, M., Nishi, T., Shibahara, T., Tanaka, T., Nakamura, Y., Isogai, T. and Sugano, S.
NEDO human cDNA sequencing project
Unpublished
2 (bases 1 to 2503)
JOURNAL Sugano, S., Suzuki, Y., Ota, T., Obayashi, M., Nishi, T., Isogai, T., Shibahara, T., Tanaka, T. and Nakamura, Y.
REFERENCE
AUTHORS Direct Submission
TITLE Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure Analysis, Human Genome Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan (E-mail: fildcna@ims.u-tokyo.ac.jp, Tel: 81-3-5449-5286, Fax: 81-3-5449-5416)
JOURNAL NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing: Department of Virology and Human Genome Center, Institute of Medical Science,

University of Tokyo (partly supported by Science and Technology

Agency).

FEATURES

BOYCE

1. 2503

CO27.1T
/organization="Homo sapiens"

```
/organelle="HOMO varrelis"
/mol type="mRNA"
```

/mol_type="tRNA"
/db_xref="taxon:9606"

/ab_xref="taxon:9606"
/gl_cof="HEDP156C3"

```
/clone="HEP15963"  
/coll_line="H10003"
```

```
/cell_line="HePG2"
```

```

/cell_type="hepatoma"
/cell_type="h1h"

```

```
/clone_lib="HEP"
```

```
/note="cloning vector
```

716. .2428

/note="unnamed protei

```
/codon_start=1
```

/protein_id="BAB15230

/db_xref="GI:10438353"
/translation="MSRLFAKPKSLCKSEPIITERTVTTLVLKRIIVTSCVGPSGRLLK
QLNGRGVGVCTSSOASLLSHLVTHPIKILITASTQNHVSFSDCGLFTALICNLN
IENVORGLPTTVITLKNKLSLSICYKSEKTCGCPVDFSTQTLCLVSRILTS
KPCAKLTRKETSEVALIRAFLLTIPNBAEGILKGLSILVPLKQGVITLSPGSI
LITMSVQDHLVPLIKSTLAKVFLCTLGSDTSGTEGVTVSVGVSLNWAQSDQ
NLNGROLIJDHVDLILCTOKVHPSLKPLFNHIIADRTGVTIMELTTCVQTQPIQ
SLGISICPNYSGVKDVCTAKGSKHFFHLINPNTICSLLLCNRTAWDELKLTQCT
ALHVLQTLKPELWALGGCTETHLAWVRHNDPSEIKDDCTQTLEQLNSAF
CSALESVSGLSHDGEELITDMKYGLHWSHADS PCVANNPDLISQCGQLNYSQBEL
NWSFLRSTCRPFVQSPQCPHEAVVSASNLTDUDCTAKLSGLQVAVETANILDLISVVI
ENKN"

ORIGIN

Query Match 97.0%: Score 2398.4: DB 9: Length 2503:

Best Local Similarity 99.5%: Pred. No. 0:

Desc Local Similarity 55.5%, Freq. NO: 0,
 Matches 2405: Conservative 0: Mismatches 11: Indels 0: Gaps 0:

1 TATGAGCCCTTCGGAACCTTGTCGAGAGACATCAAGATTTCGTTGTTATGGTCCCTTAGT 60

88 TATGAGCCTTCGGAATTGTCGAGAGACATCAAGCTTTTGGTTGTTATGCTCCATTACT 147

Q: What are some common signs and symptoms of a urinary tract infection?

A: Common signs and symptoms include frequent urination, pain or burning during urination, cloudy or bloody urine, and a strong odor.

148

149

150

151

152

153

154

155

156

157

158

159

160

161

162

163

164

165

166

167

168

169

170

171

172

173

174

175

176

177

178

179

180

181

182

183

184

185

186

187

188

189

190

191

192

193

194

195

196

197

198

199

200

201

202

203

204

205

206

207

208

209

210

211

212

213

214

215

216

217

218

219

220

221

222

223

224

225

226

227

228

229

230

231

232

233

234

235

236

237

238

239

240

241

242

243

244

245

246

247

248

249

250

251

252

253

254

255

256

257

258

259

260

261

262

263

264

265

266

267

268

269

270

271

272

273

274

275

276

277

278

279

280

281

282

283

284

285

286

287

288

289

290

291

292

293

294

295

296

297

298

299

300

301

302

303

304

305

306

307

308

309

310

311

312

313

314

315

316

317

318

319

320

321

322

323

324

325

326

327

328

329

330

331

332

333

334

335

336

337

338

339

340

341

342

343

344

345

346

347

348

349

350

351

352

353

354

355

356

357

358

359

360

361

362

363

364

365

366

367

368

369

370

371

372

373

374

375

376

377

378

379

380

381

382

383

384

385

386

387

388

389

390

391

392

393

394

395

396

397

398

399

400

401

402

403

404

405

406

407

408

409

410

411

412

413

414

415

416

417

418

419

420

421

422

423

424

425

426

427

428

429

430

431

432

433

434

435

436

437

438

439

440

441

442

443

444

445

446

447

448

449

450

451

452

453

454

455

456

457

458

459

460

461

462

463

464

465

466

467

468

469

470

471

472

473

474

475

476

477

478

479

480

481

482

483

484

485

486

487

488

489

490

491

492

493

494

495

496

497

498

499

500

501

502

503

504

505

506

507

508

509

510

511

512

513

514

515

516

517

518

519

520

521

522

523

524

525

526

527

528

529

530

531

532

533

534

535

536

537

538

539

540

541

542

543

544

545

546

547

548

549

550

551

552

553

554

555

556

557

558

559

560

561

562

563

564

565

566

567

568

569

570

571

572

573

574

575

576

577

578

579

580

581

582

583

584

585

586

587

588

589

590

591

592

593

594

595

596

597

598

599

600

601

602

603

604

605

606

607

608

609

610

611

612

613

614

615

616

617

618

619

620

621

622

623

624

625

626

627

628

629

630

631

632

633

634

635

636

637

638

639

640

641

642

643

644

645

646

647

648

649

650

651

652

653

654

655

656

657

658

659

<

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

Age Group	Percentage
18-24	10%
25-34	15%
35-44	20%
45-54	25%
55-64	20%
65-74	15%
75-84	10%
85-94	5%
95-104	5%

.....

0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99

DD

S08

C09AIAICITICTCAIAGSISIGCCATCCTTGGACACTAACCTTTTGACAASGTCISGCATTAGCACAO

T77

QY 381 ICIGTAAATAIICAGICICATAGGCATACATTAAGAGTATGTAAGATGTAAGTAAATAATAI 420

DD 448 TCIGTAAATAATTCAGTCTCATAGGCATTACATAAAGAGTAATGAAAGTGGAGTAAATAATAATA 507

421 ATTTGGAAATACIAAATTGTCAATAAATCAATTCATGCTGATAGCTTCATAAACAATGA 480

DB 508 ATTTGGAAATTACTAAATTGTCAATTAAATCAATCTAAGCTGATTAGCTTCATAAACAATTGA 567

481 ACCTTTTGAATTTATAGCCACAATGCTGCATATTCATACCTTAATTCCTAAGAATAAT 540

Db 568 ACTTTTGGATTATAGCCACAATGCTGCATATTCATACCTTTAATTCCTAAGAAATAATT 627

Qy 541 TTTAATGTTAAACGTCGATAATGCAATAAATAGAAAAATGTGGTTTACAAATAAAAAACG 600

Qy	1681	CACGCAAAATTTGGCTCCCAACATTTTTCATCTTATTTCTTAATGAACCAACAATCTG	1740
Db	1768	CACGCAAAATTTGGCTCCCAACATTTTTCATCTTATTTCTTAATGAACCAACAATCTG	1827
Qy	1741	CAGCTTGCTCTCTGCAACAGAAATGACACTGCTGGATGAGCTGAAGCTCAGCTGTCA	1800
Db	1828	CAGCTTGCTCTCTGCAACAGAAATGACACTGCTGGATGAGCTGAAGCTCAGCTGTCA	1887
Qy	1801	GACGGCACTGCATGCTCTCGCAGTTAAACACTCAAGGAACCATGGCTTTGTTGGAGGTGG	1860
Db	1888	GACGGCACTGCATGCTCTCGCAGTTAAACACTCAAGGAACCATGGCTTTGTTGGAGGTGG	1947
Qy	1861	CTGTACTGAACACTATTTGGCTGCATATATCAGACACAGACTCACAACGCCACCAAG	1920
Db	1948	CTGTACTGAACACTATTTGGCTGCATATATCAGACACAGACTCACAACGCCACCAAG	2007
Qy	1921	CATTCTCAAGATGATGAATGTACTCAACAGAACTTCAATTAATTGCTGAACATTTTG	1980
Db	2008	CATTCTCAAGATGATGAATGTACTCAACAGAACTTCAATTAATTGCTGAACATTTTG	2067
Qy	1981	CAGTGGCCTAGAATCTGTTGGTCTTTAGAACATGATGGAGGTGAATTTCTCACTGA	2040
Db	2068	CAGTGGCCTAGAATCTGTTGGTCTTTAGAACATGATGGAGGTGAATTTCTCACTGA	2127
Qy	2041	CATGAAGATGAGACACTTTGGTCAGTTTCAGGAGATTTCCCTGTGTTGCTAACTGGCC	2100
Db	2128	CATGAAGATGAGACACTTTGGTCAGTTTCAGGAGATTTCCCTGTGTTGCTAACTGGCC	2187
Qy	2101	AGATTTGCTTTACAGTGGCTGTGGATATATACATATGCGCAGAACTCAACTGGTC	2160
Db	2188	AGATTTGCTTTACAGTGGCTGTGGATATATACATATGCGCAGAACTCAACTGGTC	2247
Qy	2161	TTTCTTAAAGAACACACGCTCGTCCATTTGTGGCCACAAAGCTGCCTTCCACATGAAGCTG	2220
Db	2248	TTTCTTAAAGAACACACGCTCGTCCATTTGTGGCCACAAAGCTGCCTTCCACATGAAGCTG	2307
Qy	2221	GGCTCAGGACGACCACTTGACTTGGACTGTTTGGACTGCAAAAGCTTAGTGGCCTACAGGT	2280
Db	2308	GGCTCAGGACGACCACTTGACTTGGACTGTTTGGACTGCAAAAGCTTAGTGGCCTACAGGT	2367
Qy	2281	GGCTGTAGACAGCCAAATTTGATTTGGGATCTTTTCATATGTTATTTGAAGATAAAACTA	2340
Db	2368	GGCTGTAGACAGCCAAATTTGATTTGGGATCTTTTCATATGTTATTTGAAGATAAAACTA	2427
Qy	2341	AGAGAATAGCATGTCGTATTACAGAGAAACAAATAAACTAGTCTGTTGGCAATTTGAGA	2400
Db	2428	AGAGAATAGCATGTCGTATTACAGAGAAACAAATAAACTAGTCTGTTGGCAATTTGAAA	2487
Qy	2401	AAAAAAAAAAAAAAAAA 2416	
Db	2488	AAAAAAAAAAAAAAAAA 2503	
RESULT 7			
BD276479			
LOCUS			
143 Human Secreted Proteins.			
BD276479			
DEFINITION			
BD276479			
VERSION			
BD276479.1 GI:33086247			
KEYWORDS			
JP 2002543836-A/151.			
SOURCE			
Homo sapiens (human)			
ORGANISM			
Homo sapiens			
REFERENCE			
AUTHORS			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
TITLE			
Soppet,D.R., Florence,K.A., Duan,R.D., Bias,C.E., Komatsoulis,G.A., Young,P.E., Moore,P.A., Steven, Ruben,M. and Rosen,C.A.			
JOURNAL			
143 Human Secreted Proteins			
Patent: JP 2002543836-A 151 24-DEC-2002;			
COMMENT			
Human Genome Sciences Inc			
OS Homo sapiens			
PN JP 2002543836-A/151			
PD 24-DEC-2002			

PF	11-MAY-2000	JP 2000518448	
PR	13-MAY-1999	US 60/134068	
PI	daniel r soppet,kimberly a florence,roxanne d duan,charles e bias,		
PI	george a komatsoulis,paul e young,paul a moore,steven PI m ruben,craig a rosen		
CC			
FF	Key	Location/Qualifiers	
FT	SITE	(1906)	
FT		/note='n equals a,t,g, or c'	
FT	SITE	(1907)	
FT		/note='n equals a,t,g, or c'	
FT	SITE	(1912)	
FT		/note='n equals a,t,g, or c'	
FT	SITE	(1921)	
FT		/note='n equals a,t,g, or c'	
FEATURES			
source			
1. .1921			
/organism="Homo sapiens"			
/mol_type="genomic DNA"			
/db_xref="taxon:9606"			
ORIGIN			
Query Match			
Best Local Similarity 99.6%; Pred. No. 1.7e-298;			
Matches 1765; Conservative 2; Mismatches 5; Indels 0; Gaps 0;			
Qy	1	TATGAGCCTTCGGAACTTGTGGAGAGACTACAAAGTTTGGTTGTTATGTCCTCTTAGT	60
Db	126	TATGAGCCTTCGGAACTTGTGGAGAGACTACAAAGTTTGGTTGTTATGTCCTCTTAGT	185
Qy	61	TGGGCTCATACATTTGGGGTGGTACAGAAATCAAAAGCAGCCCTGTTTCCAAATACCTAA	120
Db	186	TGGGCTCATACATTTGGGGTGGTACAGAAATCAAAAGCAGCCCTGTTTCCAAATACCTAA	245
Qy	121	AAACGACGACATTCCTGAGCAAGATAGTCTGGGACCTTTCAAATCTTCAGAAGAGCCAAAT	180
Db	246	AAACGACGACATTCCTGAGCAAGATAGTCTGGGACCTTTCAAATCTTCAGAAGAGCCAAAT	305
Qy	181	CCAGGGGAGTAGCAGGCTTGCATCTTCAGGTAAAGAGCAGCTTTGATCTGAGCTTC	240
Db	306	CCAGGGGAGTAGTAGGCTTGCATCTTCAGGTAAAGAGCAGCTTTGATCTGAGCTTC	365
Qy	241	ATATCCAAAGAGAGATGAAAAATACACAGTTGGATTAGAAAGAACTGGCTTCTGTAGCT	300
Db	366	ATATCCAAAGAGAGATGAAAAATACACAGTTGGATTAGAAAGAACTGGCTTCTGTAGCT	425
Qy	301	GGGATATCTTTTCATAGGTGTCATCTTTGGAAACATACCTTTTGCAGAGGTCTGCAAGCAG	360
Db	426	GGGATATCTTTTCATAGGTGTCATCTTTGGAAACATACCTTTTGCAGAGGTCTGCAAGCAG	485
Qy	361	TCTGTAAATTTTCAGTCTCAAGCAAAACAAAGAGATTTGAAGAGTGAAGTAAATAAT	420
Db	486	TCTGTAAATTTTCAGTCTCAAGCAAAACAAAGAGATTTGAAGAGTGAAGTAAATAAT	545
Qy	421	ATTTGGAAATTTTAAATTTTCATTAATCAATCTATGCTGATTAGCTTTCATAAATTAATGA	480
Db	546	ATTTGGAAATTTTAAATTTTCATTAATCAATCTATGCTGATTAGCTTTCATAAATTAATGA	605
Qy	481	ACTTTTGTGATTTTATAGCCCAATGCTGCATATTCATACCTTTAAATTTCTAAAGAAATTAAT	540
Db	606	ACTTTTGTGATTTTATAGCCCAATGCTGCATATTCATACCTTTAAATTTCTAAAGAAATTAAT	665
Qy	541	TTTAATGTTTAAACGCTGATAATGCAATAATAGAAAAATGTGGTTTCAAAAATAAAACG	600
Db	666	TTTAATGTTTAAACGCTGATAATGCAATAATAGAAAAATGTGGTTTCAAAAATAAAACG	725
Qy	601	GTCTTCACCTAGTTTACCACTGGAAGTAAAGATGTCTCGTTTGGAGCTTAAGAGCCATCAT	660
Db	726	GTCTTCACCTAGTTTACCACTGGAAGTAAAGATGTCTCGTTTGGAGCTTAAGAGCCATCAT	785
Qy	661	GTGTAGAGGTGAACCACTGCAAACTGAGAGAGTCAAGCAACACTTTCTGTCTTGAAG	720

Db 786 GTGTAAAGAGTGAACCACTGACAACTGAGAGAGTCAGGACCAACACATTTTCTGTCTTGAAGAG 845
QY 721 AATTGTAAACATCATGTATGGCCCTCAGTAGGCTGAAGCAGCTGCACAAATGGCTTTGG 780
Db 846 AATTGTAAACATCATGTATGGCCCTCAGTAGGCTGAAGCAGCTGCACAAATGGCTTTGG 905
QY 781 AGTTACGTTGTGACAACTCAGAGTCCCTCAGCTCTGCTCAGTCACTTTTGGTACACA 840
Db 906 AGTTACGTTGTGACAACTCAGAGTCCCTCAGCTCTGCTCAGTCACTTTTGGTACACA 965
QY 841 TCCATTTTAAAGATCCTGACAGCTCCATACAGAAATCATGTGTCAAGCTTCAGTGATTG 900
Db 966 TCCATTTTAAAGATCCTGACAGCTCCATACAGAAATCATGTGTCAAGCTTCAGTGATTG 1025
QY 901 TGGCTTATTCAAGCTATTCTTTGCTGCAACCTGATTGAAATGTTCAAGATTAGGCTT 960
Db 1026 TGGCTTATTCAAGCTATTCTTTGCTGCAACCTGATTGAAATGTTCAAGATTAGGCTT 1085
QY 961 GACACCCACACCTGTCTATTAGATTAAATAAACAATCTTTTGAGTCTTTGGCATCAGTTATCT 1020
Db 1086 GACACCCACACCTGTCTATTAGATTAAATAAACAATCTTTTGAGTCTTTGGCATCAGTTATCT 1145
QY 1021 CAACTCTGATACCTGTGGTTGTCGAATCCAGTGGACTTTTAGTAGTACTCAGATCCTCT 1080
Db 1146 CAACTCTGATACCTGTGGTTGTCGAATCCAGTGGACTTTTAGTAGTACTCAGATCCTCT 1205
QY 1081 TTGTTTGGTGGTAGTATATTAAAGATAAACCCTGCTGTATGCTCACAGAAAGAAAC 1140
Db 1206 TTGTTTGGTGGTAGTATATTAAAGATAAACCCTGCTGTATGCTCACAGAAAGAAAC 1265
QY 1141 AGAGCATGTCAGTGTCTTGATCTGAGAGCTTTTGGCTTTTACAAATCCAGAAATCCTCA 1200
Db 1266 AGAGCATGTCAGTGTCTTGATCTGAGAGCTTTTGGCTTTTACAAATCCAGAAATCCTCA 1325
QY 1201 AGSCCAATCATTTTAAAGAAAGTTTAAATGTGACCTTTTAAAGAGTCAAGAGTTATAGA 1260
Db 1326 AGSCCAATCATTTTAAAGAAAGTTTAAATGTGACCTTTTAAAGAGTCAAGAGTTATAGA 1385
QY 1261 TTCCACTGTATTTACCTGGGATCTCAATGAAATGTGAGAGTTCAATTAATGAGGCTATT 1320
Db 1386 TTCCACTGTATTTACCTGGGATCTCAATGAAATGTGAGAGTTCAATTAATGAGGCTATT 1445
QY 1321 ACCTATCAAAAAATCAACTGSCCTCAAGGTGGCACTCTTTTGTACAACTTTATCCGGAGA 1380
Db 1446 ACCTATCAAAAAATCAACTGSCCTCAAGGTGGCACTCTTTTGTACAACTTTATCCGGAGA 1505
QY 1381 CACTTCTGACATGGAAGAGAACTGTGTGTGTGTCAGTTATGGGTTTCTCTTGAATAATGC 1440
Db 1506 CACTTCTGACATGGAAGAGAACTGTGTGTGTGTCAGTTATGGGTTTCTCTTGAATAATGC 1565
QY 1441 AGTCTTGGACCACTGCTTAACTAGGAAGGAGCTTAATCAGTGACCACTGATCTTGT 1500
Db 1566 AGTCTTGGACCACTGCTTAACTAGGAAGGAGCTTAATCAGTGACCACTGATCTTGT 1625
QY 1501 CCTGTGCCAAAAAGTTATACATCCATCTTTGAAGCAGTTTCTCAATGATCATCGTATTAT 1560
Db 1626 CCTGTGCCAAAAAGTTATACATCCATCTTTGAAGCAGTTTCTCAATGATCATCGTATTAT 1685
QY 1561 TGCATAGACAGAAATGGAGTACTCTGATGGAAACCCCTGACTTAAATGACAGGAAACACA 1620
Db 1686 TGCATAGACAGAAATGGAGTACTCTGATGGAAACCCCTGACTTAAATGACAGGAAACACA 1745
QY 1621 GCCTATTGGATCCCTAGGCTCAATATGCTCTAATAGTTATGGAAGTGTGAAGATGTGTG 1680
Db 1746 GCCTATTGGATCCCTAGGCTCAATATGCTCTAATAGTTATGGAAGTGTGAAGATGTGTG 1805
QY 1681 CACTGCAAAATTTGGCTCCCAACATTTTTCATCTTATTCTTAAATGAAGCAACAATCTG 1740
Db 1806 CACTGCAAAATTTGGCTCCCAACATTTTTCATCTTATTCTTAAATGAAGCAACAATCTG 1865
QY 1741 CAGCTTGTCTCTCTGCAACAGAAATGACACTG 1772
Db 1866 CAGCTTGTCTCTCTGCAACAGAAATGAGGGTG 1897

RESULT 8
CQ729071
LOCUS 1713 bp DNA linear PAT 03-FEB-2004
Sequence 15005 from Patent WO02068579.
CQ729071
ACCESSION
CQ729071.1 GI:42299608
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS
TITLE
Venter, C.J., Adams, M.C., Li, P.W. and Myers, E.W.
Kits, such as nucleic acid arrays, comprising a majority of
humanexons or transcripts, for detecting expression and other uses
thereof
JOURNAL
Patent: WO 02068579-A 15005 06-SEP-2002;
PE Corporation (NY) (US)
FEATURES
Location/Qualifiers
source
1..1713
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 69.1%; Score 1708.2; DB 6; Length 1713;
Best Local Similarity 99.8%; Pred. No. 7.3e-289;
Matches 1710; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 629 ATGCTCTCGTTTGGAGCTAAGAGCCATCATTTGTGTAGAGTGAACCACTGCACACTGAG 688
Db 1 ATGCTCTCGTTTGGAGCTAAGAGCCATCATTTGTGTAGAGTGAACCACTGCACACTGAG 60
QY 689 AGAGTCAGGACCAACACTTTCTCTCTTTGAAAGAAATTTAAACATCATGCTATGGCCCTCA 748
Db 61 AGAGTCAGGACCAACACTTTCTCTCTTTGAAAGAAATTTAAACATCATGCTATGGCCCTCA 120
QY 749 GGTAGCTGAAGCAGCTGCACAAATGGCTTTGGAGTTAGCTGTGTACAACTGCACAGTCC 808
Db 121 GGTAGCTGAAGCAGCTGCACAAATGGCTTTGGAGTTAGCTGTGTACAACTGCACAGTCC 180
QY 809 TCAGCTCTGCTCAGTCACCTTTTGGTCAACATCCATTTTAAAGATCCTGCAGAGCTCC 868
Db 181 TCAGCTCTGCTCAGTCACCTTTTGGTCAACATCCATTTTAAAGATCCTGCAGAGCTCC 240
QY 869 ATACAGAAATCATGTGTCAAGCTTTCAGTGATTTGGCTTATTCACAGCTATTCTTTGCTGC 928
Db 241 ATACAGAAATCATGTGTCAAGCTTTCAGTGATTTGGCTTATTCACAGCTATTCTTTGCTGC 300
QY 929 AACCTGATTGAAATGTTTCAGAGATTAGGCTTGACACCCACCATGTGTCAATTAGATTAAAT 988
Db 301 AACCTGATTGAAATGTTTCAGAGATTAGGCTTTGACACCCACCATGTGTCAATTAGATTAAAT 360
QY 989 AAGCATCTTTTTCAGCTTTTGGCATCAGTTATCTCAAGTCTGAGACCTGTGTGCGAATC 1048
Db 361 AAGCATCTTTTTCAGCTTTTGGCATCAGTTATCTCAAGTCTGAGACCTGTGTGCGAATC 420
QY 1049 CCAGTGGACTTTTAGTAGTACTCAGATCCTCTTTGTTGGTGGTAGTAGTATTTAAACAAGT 1108
Db 421 CCAGTGGACTTTTAGTAGTACTCAGATCCTCTTTGTTGGTGGTAGTAGTATTTAAACAAGT 480
QY 1109 AAACCTGCTGTATGCTCACAGAAAGAAACAGAGCATGTCAAGTCTTTGATTCCTGAGA 1168
Db 481 AAACCTGCTGTATGCTCACAGAAAGAAACAGAGCATGTCAAGTCTTTGATTCCTGAGA 540
QY 1169 GCCTTTTTCGCTTACAAATTCAGAAATGCTGAAGCCACATCATTTTGAAGAGAGTTTA 1228
Db 541 GCCTTTTTCGCTTACAAATTCAGAAATGCTGAAGCCACATCATTTTGAAGAGAGTTTA 600
QY 1229 ATTGTACCTTTAAAGAGTCAAGAGTTATAGATTCCACTGTATTACCTGGGATCTCATTT 1288

Db	601	ATTGTACCTTTAAAGGTCAAAGAGTTATAGATTCCACTGTATTCCTGGATACCTACTT	660
Qy	1289	GAATGTGAGAGTTCAATTAATGAGGCTATTACCTATCAAAAAATCAACTGCGCTCAAG	1348
Db	661	GAATGTGAGAGTTCAATTAATGAGGCTATTACCTATCAAAAAATCAACTGCGCTCAAG	720
Qy	1349	GTGGCACTCTTTTGTACAACTTTATCGGAGACACTTCTGACACTCGAGAAGAACTGTG	1408
Db	721	GTGGCACTCTTTTGTACAACTTTATCGGAGACACTTCTGACACTCGAGAAGAACTGTG	780
Qy	1409	GTGGTCAGTTATGGGGTTCTCTTGAATAATGCACTCTTGACACAGCTGCTTAACCTAGGA	1468
Db	781	GTGGTCAGTTATGGGGTTCTCTTGAATAATGCACTCTTGACACAGCTGCTTAACCTAGGA	840
Qy	1469	AGCAGCTAATCAGTGACCACTAGATCTTGTCTGTGCCAAAAGTTATACATCCATCT	1528
Db	841	AGCAGCTAATCAGTGACCACTAGATCTTGTCTGTGCCAAAAGTTATACATCCATCT	900
Qy	1529	TTGAAGCAGTTTCTCAATATGCAATGATATTTATTTGCCATAGACAGAATGGAGTGACTCTG	1588
Db	901	TTGAAGCAGTTTCTCAATATGCAATGATATTTATTTGCCATAGACAGAATGGAGTGACTCTG	960
Qy	1589	ATGGACCCCTGACTAAATGACAGGAACACAGCCTATTGGATCCCTAGGCTCAATATGT	1648
Db	961	ATGGACCCCTGACTAAATGACAGGAACACAGCCTATTGGATCCCTAGGCTCAATATGT	1020
Qy	1649	CCTAATAGTTATGGAAGTGTGAAGATGTGCACTGCAAAATTTGGCTCCCAACATTTT	1708
Db	1021	CCTAATAGTTATGGAAGTGTGAAGATGTGCACTGCAAAATTTGGCTCCCAACATTTT	1080
Qy	1709	TTTCATCTTATTCCTAATGAAGCAACAATCTGCAGCTTGCTCTCTGCAACGAAATGAC	1768
Db	1081	TTTCATCTTATTCCTAATGAAGCAACAATCTGCAGCTTGCTCTCTGCAACGAAATGAC	1140
Qy	1769	ACTGCTGGATGAGCTGAAGCTCAGTGTGACAGCGCACTGATGCTCTGCAAGTAAACA	1828
Db	1141	ACTGCTGGATGAGCTGAAGCTCAGTGTGACAGCGCACTGATGCTCTGCAAGTAAACA	1200
Qy	1829	CTCAAGAACCTAGGCTTTGTTGGAGTGGCTGACTGAACCTATTTGGCTGCATAT	1888
Db	1201	CTCAAGAACCTAGGCTTTGTTGGAGTGGCTGACTGAACCTATTTGGCTGCATAT	1260
Qy	1889	ATCAGACACAAGACTCACAACCGCCAGAAAGCAATTTCTCAAGATGATGAATGACTCAA	1948
Db	1261	ATCAGACACAAGACTCACAACCGCCAGAAAGCAATTTCTCAAGATGATGAATGACTCAA	1320
Qy	1949	ACGAACCTCAATTAATTTGCTGAAGCAATTTGCACTGAGTGCCTAGAACTCTGTTGGCTCT	2008
Db	1321	ACGAACCTCAATTAATTTGCTGAAGCAATTTGCACTGAGTGCCTAGAACTCTGTTGGCTCT	1380
Qy	2009	TTAGAACATGATGAGAGTGAATTTCTCACTGACATGAAGTATGACACACCTTTGGTCAGTT	2068
Db	1381	TTAGAACATGATGAGAGTGAATTTCTCACTGACATGAAGTATGACACACCTTTGGTCAGTT	1440
Qy	2069	CAGGCAGATTCTCCCTGTGTGCTAACTGCGCAAGATTTGCTTTCAAGTGTGCTGTGGA	2128
Db	1441	CAGGCAGATTCTCCCTGTGTGCTAACTGCGCAAGATTTGCTTTCAAGTGTGCTGTGGA	1500
Qy	2129	TTATACAATAGCAGAGAACTCAACTGCTCTTTCTTAAGAGACACAGTGTGCTCAATTT	2188
Db	1501	TTATACAATAGCAGAGAACTCAACTGCTCTTTCTTAAGAGACACAGTGTGCTCAATTT	1560
Qy	2189	GTGCCACAAGCTGCTTCCACATGAAGTGTGGGCTCAGCCAGCAACCTGACCTGGAC	2248
Db	1561	GTGCCACAAGCTGCTTCCACATGAAGTGTGGGCTCAGCCAGCAACCTGACCTGGAC	1620
Qy	2249	TGTTTGACTGCAAGCTTAGTGGCTTACAGGTGGCTGTAGACAGCCAAATTTGATTTGG	2308
Db	1621	TGTTTGACTGCAAGCTTAGTGGCTTACAGGTGGCTGTAGACAGCCAAATTTGATTTGG	1680
Qy	2309	GATCTTTTCAATGTTTATGAAGATAAAAACTAA	2341
Db	1681	GATCTTTTCAATGTTTATGAAGATAAAAACTAA	1713

RESULT 9

BC083863

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

BC083863

Rattus norvegicus cDNA clone MGC:95109 IMAGE:7127425, complete cds.

BC083863.1

GI:54035334

MGC.

Rattus norvegicus (Norway rat)

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE

1 (bases 1 to 2972)

STRAUSBERG R.L., FEINGOLD E.A., GROUSE L.H., DERGE J.G., KLAUSNER R.D., COLLINS F.S., WAGNER L., SCHAFFNER C.M., SCHULER G.D., ALTSCHUL S.F., ZEEBERG B., BUETOW K.H., SCHAEFER C.F., BHAT N.K., HOPKINS R.F., JORDAN H., MOORE T., MAX S.I., WANG J., HSIEH F., DIATCHENKO L., MARUSINA K., FARMER A.A., RUBIN G.M., HONG L., STAPLETON M., SOARES M.B., BONALDO M.F., CAAVAANT T.L., SCHEETZ T.E., BROWNSTEIN M.J., USLIN T.B., TOSHIYUKI S., CARINCI P., PRANGE C., RAHA S., LOQUELLANO N.A., PETERS G.J., ABRAMSON R.D., MULLAHSY S.J., BOSAK S.A., MCEWAN P.J., MCKERNAN K.J., MALEK J.A., GUNARATNE P.H., RICHARDS S., WORLEY K.C., HALE S., GARCIA A.M., GAY L.J., HULYK S.W., VILLALON D.K., MUZNY D.M., SODERGREN E.J., LU X., GIBBS R.A., FAHEY J., HELTON E., KETTEMAN M., MADAN A., RODRIGUES S., SANCHEZ A., WHITING M., MADAN A., YOUNG A.C., SHEVCHENKO Y., BOUFFARD G.G., BLAKESLEY R.W., TOUCHMAN J.W., GREEN E.D., DICKSON M.C., RODRIGUEZ A.C., GRIMWOOD J., SCHMUTZ J., MYERS R.M., BUTTERFIELD Y.S., KRZYWINSKI M.I., SKALSKA U., SMALLUS D.E., SCHNERCH A., SCHEIN J.E., JONES S.J. and MARRA M.A.

Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences

Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

12477932

REFERENCE

2 (bases 1 to 2972)

Director MGC Project.

Direct Submission

Submitted (01-OCT-2004) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: cgapbs-remail.nih.gov

Tissue Procurement: Howard Jacobs

cDNA Library Preparation: Express Genomics

DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)

Center, Stanford University School of Medicine, Stanford, CA 94305

Web site: <http://www-shgc.stanford.edu>

Contact: (Dickson, Mark) mcd@paxil.stanford.edu

Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

FEATURES

source

1..2972

/organism="Rattus norvegicus"

/mol_type="mRNA"

/db_xref="taxon:10116"

/clone="MGC:95109 IMAGE:7127425"

/tissue_type="Kidney, rat (Brown Norway)"

/clone_lib="NIH MGC_236"

/lab_host="DH10B"

/note="Vector: pExpress1"

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRAP Plate: 184 Row: a Column: 6

This clone was selected for full length sequencing because it passed the following selection criteria: Similarity but not identity to protein.

Location/Qualifiers

Db 480 TCTGTGAGCTCTGAGCCAGAGGACACAAAAGAAATATTGAAGAATGAAGCAAAATAAAG 539
Qy 421 ATTGTGAATTAATAATTGTCATTAATAATCATTTCTATGCTGATAGCTTCATAAACATTTGA 480
Db 540 ACTTTGGATTAATTA--TGTCAATTAATAATTAATCTATGCTGGTTATTGCTGAAACCACTT 597
Qy 481 ACTTTTGTGATTTTATAGCCACCAATGCTGCATATTCAATGCTTTAAATTTCTTAAGAAATAATT 540
Db 598 GATTTTGTAGCTGGT-----AATGCTGCATGTCAGAC--TCATACCTGGAGACAAGTC 649
Qy 541 TTTAATGTTTAAACGTCGTAATAATGCAATAAATAGAAAAATGTGTTTTACAAAAATAAAACG 600
Db 650 TTTAATGTTTCAACACATGACAGTGCATTAATAATGAAGCCCATATGAACAAATTAAGCT 709
Qy 601 GTCTTCACTAGTTTACACCTGAGAGTAAGATGTCCTGTTTGGTGGAGCTTAAGAACCAATCAT 660
Db 710 GCCTTCAACAAGTTA-CACCTGAAGCAGGATGTCCTCGTTTAGAAGCTTAAGAACCAATCGTT 768
Qy 661 GTGTAGAGTGAACCACTGACAACTGAGAGAGTGCAGGACCACACTTTTCTGCTTGAAGAAG 720
Db 769 GTGTAAACTGAACCAATTGACAAAGTGAGAAAGTCAGGTCACACTTTTCTGCTTGAAGAAG 828
Qy 721 AATTGTAAACATCATGCTATGGCCCTCAGGTAGGCTGAAGCAGCTGCACAATGGCTTTGG 780
Db 829 AGTCATAGCTCATGCTATGGCCCTTCAGGGAGCTGAAGCAGCTACACAATGGATTGG 888
Qy 781 AGGTTAGGTGTGTAACACTGACAGTCCCTCAGCTCTGCTCAGTCACTTTTGGTGCACACA 840
Db 889 AGGTTGTGTGTACACAACCTGACAGTCCCTCAGGCCCTGCTCGAATCCCTTCAGTCAACCA 948
Qy 841 TCCCATTTTAAAGATCCTGACAGCCTCCATACAGAATCATGTGTCGAAGCTTCAGTGATG 900
Db 949 TCCCGTATTAAAGATCCTGACAGTCCCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 1008
Qy 901 TGCGTTATTACAGCTATTCTTTGCTGCAACCTGATTTGAAATGTTTACAGATTAAGCTT 960
Db 1009 CGGCTTATTACAGCCATTTCTGTGCTGCACTTTGATTGAATTAATTTCAAAGACTAGATT 1068
Qy 961 GACACCACACCTGCTATTAGATTAAATAAACAATCTTTTGGTCTTTGGATCAGTTATCT 1020
Db 1069 GACACCCGCACTGCTATTAAATAAATAAATACCTCTTTGAGTCTCTGTACCAGTTATCT 1128
Qy 1021 CAAGTCTGATACCTGCTGTTGTCGAATCCAGTGGACTTTTGTAGTACTCAGATCCTCT 1080
Db 1129 CAAGTCTGAAGCTGTAGTTGTCGAATCCAGTTGACTTCAGAAATGACATACCTTCT 1188
Qy 1081 TTGTTTGGTGGTAGTATATTAAACAAGTAAACCTGCTGTATGCTCACACAGAAAGAAAC 1140
Db 1189 CAGCTTGGTACACAGTATCTTAAACAAGCAAAACAGCTGTATGCTCACACAGAAAGAAAC 1248
Qy 1141 AGAGCATGTCAGTCTTTGATCCTGAGAGCTTTTGTGCTTACAAATTCAGAAATGCTGA 1200
Db 1249 AGATCATAGTGGTCTTGAATTTGAAGCTTTTGTACTTACAAATTCAGAAATGACAGA 1308
Qy 1201 AGGCCACATCATTTTAGGAAAGAGTTTAAATTGTACCTTTTAAAGGTCAAAGAGTTATAGA 1260
Db 1309 AGAACGAATGGTTTTAGGAGAGAGTATTAATGTTCCCTTAAAGGGCAAGAGTTACAGA 1368
Qy 1261 TTCCTGTTATTAACCTGGGATCTCTATGAAATGTCAGAAATTCATTAATGAGGCTATT 1320
Db 1369 TTCTACTGTTATTAACCTGGGACTCTCATTTGAAGCATCAGAAGTTTCAATTAAGGAGGTTAT 1428
Qy 1321 ACCTATCAAAAATCAATGCCCTCAAGGTGGGACTTTTGTGACAACTTTATCCGGAGA 1380
Db 1429 ACCACTCAAAAGGCAAGTGGCTTAAAGGTGGGACTTCTCTGTACATCTTTATCTGGAGA 1488
Qy 1381 CACTTCTGACATCTGGAGAGAGAACTGTGGTGGTCAAGTTATGGGTTTCTCTTGAATAATGC 1440
Db 1489 TTTTCTAATGCTGGAGAGAGAGTGTGGTGGGCCAATACCAAGTCTCCCTTGAAGATGC 1548
Qy 1441 AGCTTTGGACCACTGCTTAAACCTAGGAAGGCAAGCTTAATCAGTGACCACTAGATCTTGT 1500
Db 1549 AGTTTTAGACAGTGTGCTTAATCTAGGAAGGCGACTAGTCACTGATCATGTAGACCTTGT 1608

Qy 1501 CCTGTGCCAAAAGTTATACATCCATCTTTGAGCAGTTTCTCAATATGCATCGTATTAT 1560
Db 1609 CCTGTGCCAAAAGTTATACATCCATCTTTGAGCAGTTTCTCAATATGCATCGTATTAT 1668
Qy 1561 TGCCATAGACAGAAATGGAGTGACCTCTGATGGAAACCCCTGACTAAATAAGACAGAAACACA 1620
Db 1669 GGCCATCGACAGAGTTGGGTGACTCTGATGGAATCTCTAGCAAAAGTGACAGAGGACAA 1728
Qy 1621 GCCTATTGGAATCCCTPAGGCTCAATATGCTTAATAGTTATGGAAAGTGTGAAAGATGTGTG 1680
Db 1729 GCCTATTGGTTCTCTAAAACCCCAATAGTTTCTACTACTTATGGAAGTGTGAAAGATGTGTG 1788
Qy 1681 CACTTCGAAAATTTGGCTCCCAACATTTTTTTCATCTTATCTCTAATAGCAACAATCTG 1740
Db 1789 CTCTCGAAGATTTGGCTCCAAACATTTTTTTCATCTTCTTCTAATAGGCGCACTGTCTG 1848
Qy 1741 CAGCTTGTCTTCTCGCAACAGAAATGACATGCTCCCTGGGATGAGCTGAAGCTCACGTGTCA 1800
Db 1849 CACCTTGTCTTCTCGCAGCAGAAATGACACCGCTGGGAGGAGCTGAAGCTCACATGTCA 1908
Qy 1801 GAGCGCACTGCAATGCTCTGCAAGTTAACTCAAGGAAACATGGGCTTTTGTGGGAGGTGG 1860
Db 1909 AACAGCAATGCAGCTCTTGCAAGTTAACTCAAGGAAACCGTGGGTTTATTGGGAGGTGG 1968
Qy 1861 CTGTACTGAAACTCATTTGGCTGCAATATACAGACACAGACTCACACGACCCAGAAAG 1920
Db 1969 CTGTACAGAAACACACTTGGCTGCAATATGTCAGACACAGGTTTCATCAGAGGCAAGAGC 2028
Qy 1921 CATTTCTCAAGATGATGATGATGATCTCAAAACAGAACTTCAATTTAAATGCTGAAGCATTTTG 1980
Db 2029 TATTGTCAGAGATGATGGGTGACTCAGGCAAAAGCTGCATGTTGCTGCTGAAGCATTTTG 2088
Qy 1981 CAGTCCCTTAGAATCTGTTGTTGGCTCTTTAGAAACATGATGAGAGTGAAAATCTCACTGA 2040
Db 2089 CAGTCTCTGGAGTCCGTTGCTGGCTCTTTGGAAACATGATGGTGGTGAATCCTCATTTGA 2148
Qy 2041 CATGAAGTATGACACCTTTGGTCAAGTTCAGGCAAGTTCTCCCTGTTGCTTAACCTGGCC 2100
Db 2149 CAGGAAGTATGACACCTTTGGTCTGCTCAAGCAGATCTGCTCTGTTGGTAACTGGTC 2208
Qy 2101 AGATTGCTTTTCAAGTGTGGCTGTGGATTATACAAATAGCCAGGAAAGAACTCAACTGGTC 2160
Db 2209 AGATAGCTGTGTCAGGCTGTGGCTGTGGTTTGTACAAACAGCAGGAAAGCTCAGCTGGTC 2268
Qy 2161 TTTCTTAAGAAAGCAACGCTGCTCAATTTGTGTCACAAAGCTGCCCTTCCACATGAAGCTGT 2220
Db 2269 TGTCTTAAGAAAGTACTTATCATCTCTTTTGCACCAAAACCTGCCCTTCCACAGGCACTTT 2328
Qy 2221 GGGCTCAGCCAGCAACCTGACCTTGGACTGTTTGGACTGCAAGCTTGTAGGCTTACAGGT 2280
Db 2329 GGGCTCAGCCAGTAACTGACTGTGGACTGCTTCACTGCAAGCTGAGTGGCTTACAGT 2388
Qy 2281 GGCTGTAGAGACAGCAATTTGATTGGGATCTTTTATATGTTTATGAAATATAAAACCTA 2340
Db 2389 GGCTGTAGAGACAGCAATTTGATTGAACTTTTAGAATCTTTCATATGTCATTGAAGATAAAACCTA 2448
Qy 2341 AGAAGATAGCATG 2353
Db 2449 ACATAAGAGAATG 2461

RESULT 11
AF254074
LOCUS
DEFINITION
Mus musculus MKKS protein (Mkks)
ACCESSION
AF254074.1
VERSION
AF254074.1
KEYWORDS
Mus musculus (house mouse)
SOURCE
ORGANISM

AF254074 2699 bp mRNA linear ROD 02-JUN-2000
Mus musculus MKKS protein (Mkks) mRNA, complete cds.
AF254074.1 GI:8164187
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE	ORIGIN
1 (bases 1 to 2699)	Query Match 62.5%; Score 1545.8; DB 10; Length 2699;
Stone,D.L., Slavotinek,A., Bouffard,M., Banerjee-Basu,S.,	Best Local Similarity 80.3%; Pred. No. 1.6e-260;
Baxevanis,A.D., Barr,M. and Biesecker,L.G.	Matches 1890; Conservative 0; Mismatches 447; Indels 16; Gaps 6;
Mutation of a gene encoding a putative chaperonin causes	
McKusick-Kaufman syndrome	
Nat. Genet. 25 (1), 79-82 (2000)	
20264373	1 TATCAGCGCTTCGGAACCTGTGCGAGAGCTACAAAGTTTGTGTGTATGTGTCCTTTAGT 60
10802661	157 TATCAGCTTTTCTGAGAGAGCTACAAAGTTTGTGTGTATGTGTCCTTTAGT 216
REFERENCE	61 TGGGCTCATACATTTGGGGTGGTACAGAAATCAAAGCAGCCCTGTTTTCCAAATACCTAA 120
AUTHORS	217 TGGGTTTCATCATTTGGGGTGGCAGAGTCAAAGTCAAAGCAGCCCTGTTTCCAAGTTCTT 274
TITLE	121 AAACGACGACATCTCTGAGCAGATAGTCTGGGACTTTCAAATCTTTCAGAAAGCCAAAT 180
MEDLINE	275 -AAGGACGATACCACTGAGCTGGGATAGTCTGGGACTTTCAAAGTCAAAGCAGCCAAAC 333
PUBMED	181 CCAGGGAAGTAGCAGGCTTCGCAATCTTCAGTAAAGCAGCTTTGAAATCTCAGCTTC 240
AUTHORS	334 CTAGGAGAGTAGCAGCTTCGAGTTTTCAG--AAAGAGCACTTTGAAATCTCAGCTTC 391
TITLE	241 ATATCGAAAGAGAGATGAAAAATACCAAGTTGGATTGAAAGAACTTGGCTTCTTTGTAGCT 300
JOURNAL	392 ATATAAAGAAAGAAATGAAAAACCCAGTTGGATTGAAAGAACTTGGCTTCTTTGTGCT 451
Submitted (10-APR-2000) NHGRI, NIH, 49 Convent Drive, Bethesda, MD	301 GGGATATCTTTCATAGGTGTCATCTTTGGAAATATCTTTTTCGAGAGGTCTGCAAGCAG 360
20892, USA	452 GGGGTGACTTTTCATAGGCGTCCATCTTTGGAAATATCTTTTACAGAGAGTTGCAAAAGAG 511
FEATURES	361 TCTGTAAATTTTCAGTCTCAAAGCAACAAAGAGTATTGAGAGTCTGAGTGAATTAATTAAT 420
source	512 TCTGTGAGGTCTGAGGCGAGGAGCAAAAGAAATATTGAAAGAAATGAAGCAAAATTAAG 571
1. .2699	421 ATTTGGAATTAATAATTTGTCAATTAATCATTTCTATGCTGAATAGCTTTCATTAATTAAG 480
/organism="Mus musculus"	
/mol_type="mRNA"	
/db_xref="taxon:10090"	
1. .2699	
/gene="Mkks"	
/gene="Mkks"	
/note="putative chaperonin"	
/codon_start=1	
/product="MKKS protein"	
/protein_id="AAF3965.1"	
/db_xref="GI:8164188"	
/translation="MSRLAKKPSLCKTEPTSEKVRSTLSVLKGVASLCYGPSGRLLK	
OLHNLGSCVYTTSSQSLALNLSVTHPLKLTSSVQNHVSVCSDELGFTAILCCNN	
IENQRDLDPATALKUNKYLLSLCTSLKSEACSCRPVDFRSTRLTSLVSHILTS	
KPACMLTREDTHGALILKAFLLTIPSTEEBVMVGLKSIIVPLKQGVDTSTVPLG	
LIASEVQLRSLDTQKASGRVALFCTSLSGDFSNAGEGVVVAHYQVSLNAVLEQI	
LNGLRLVITDHVLDVQKVHPSLRQPFSSRHVMAIDRVGVITLMESLSKVGTATPG	
SLNP1VSTTVSVKDVCSARGSKHFFPHLLPNEATVCTLLLCRSNDTAWELKLTCT	
AMHVLOITIKEPWLLGGGCTETHLAAYVRHKVHHEAAIVRDDGCTCAQLHVAARH	
CSALESVAGSLHGGGELLIDTKYGHLSQDASVGNWSDITLSRCCGGLYNSQEEI	
SWSVLRSTYHFPAFOTCLPQAAALGSASNLTVDCFTAKLSGLQAVETANLILDLISYV	
EDKN"	

572	Db	ACTTTGGATTTATTAA--TGTCAATTAATTAATCTATGCTGGTTATTGTCTGAAACCACTTT	629
481	Qy	ACTTTTTTGATTTTATAGCCACCAATGCTGCATATTTCAATCTTTTAATTCCTTAAGAAATAATT	540
630	Db	GAATTTGTAGTGGT-----AATGCTGCATGTGCAGAC--TCATTAACCTGGAGCCAGATC	681
541	Qy	TTTTAAATGTTAAAAACGTGATAATGCAATAAATAGAAAAATGTGTTTTACAAAAATAAAACG	600
682	Db	TTTTAAATGTTCAAAACACATGACAGTGCATAATAAATGAAAGCCATATGAACAAACTAAAGCT	741
601	Qy	GTCTTCACCTAGTTACACCTGAAAGTGAAGATGTCTCTGTTTGGAAAGCTTAAGAGCCATCAT	660
742	Db	GCCTTCACCAAGTTA-CACCTGAAGCAGGATGTCTCGTTTAGAAGCTTAAGAAGCCATCGTT	800
661	Qy	GTCTTAAGAGTGAACCACTGACAACCTCAGAGAGTCAAGSACCACACTTTCTGTCTTTGAAAG	720
801	Db	GTGTAAACCTGAACCAATGACAAGTGAGAAAGTCAAGTCCACACTTTCTGTCTTTGAAAGG	860
721	Qy	AAATTGTAACATCATGCTATGGCCCTTCAGGTAGGCTGAAGCAGCTGCACAATGGCTTTGG	780
861	Db	AGTCATAGCTCATGCTATGGCCCTTCAGGAGGCTGAAGCAGCTACACATGATTTGGG	920
781	Qy	AGGTTACGTGTGTACAACCTCAGCTCCTCAGCTCTGCTCGTCACTGTCACCTTTTGGTCAACA	840
921	Db	AGGTTGTGTGTACACAACCTTCAGTCTCAGGCTCTCGAGCCCTGCTTCGAAAGCTTTTCAGTCA	980
841	Qy	TCCCAATTTAAAGATCCTGACAGCCTCCATACAGAAATCATGTGTCAAGCTTCAGTCAATG	900
981	Db	TCCCGTATTAAAGATCCTTAACGTCATCCGCTGCAGAAATCAGTGTCTGCTCTTCAGTGA	1040
901	Qy	TGCGTTATTACACAGCTATTCTTTTGTGCAACCTGATTTGAAATGTTTCAAGATTTAGGCT	960
1041	Db	CGGCTTTATTACAGCCATTCTGTGCTGCAACTTGATTGAAATATTCAAGAGCTAGATT	1100
961	Qy	GACACCCACACGTGTCATTAGATTAAATAAAACATCTTTTGAAGTCTTTGTCATCAGTTATCT	1020
1101	Db	GACACCCGCAACTGCTATTAAATTAATAAATACCTCTTGAAGTCTCTGTACCAAGTTATCT	1160
1021	Qy	CAAGTCTGATACCTGTGGTTGTGCATCCAGCTGGAGCTTTTAGTGTACTTCAGATCCCTCT	1080
1161	Db	CAAGTCTGAAGCCTGTAGTTGTGCAATCCAGTTGACTTCAGAAGTACACATACCTTCTCT	1220
1081	Qy	TTGTTTGGTGGGTAGTATATTAAACAGTAAACCTGCTGTATGCTCACACAGAAAGAAAC	1140
1221	Db	CAGCTTGGTACAGATCTTTAAACAGCAAAACAGCTGTATGCTCACCAAGAAAGAAAC	1280
1141	Qy	AGAGCATGTCAGTGCTTTGATCCTGAGAGCCTTTTGTCTTACAATTCAGAAATGCTGA	1200
1281	Db	AGATCACATAGGTGCTTTGATTTTGAAGCTTTTTTACITTAACAATTCAGAAAGTACAGA	1340
1201	Qy	AGGCCACATCATTTTAGGAAAGATTAAATGTGACTTTTAAAGGTCAAAGATTATAGA	1260
1341	Db	AGAACGAATGGTTTTAGGGAAGAGTATAATGTTCTTAAAGGGCCAAAGAGTTACAGA	1400
1261	Qy	TTCCACTGTATTACCTGGGATCTCAATCAAAATGTCAGAAAGTTCAATTAATGAGGCTATT	1320
1401	Db	TTCTACTGTATTACCTGGACTACTCATGAGCATCAGAAGTTCAATTAAGGAGGTTATT	1460
1321	Qy	ACCTATCAAAAAATCAACTGCCCTCAAGGTGGCACTCTTTTGTACAACTTTATCCGGAGA	1380
1461	Db	ACCCACTCAAAAGGCAAGTGGCTTAGGGTGGCACTCTTCTGTACATCTTTATCTCGAGA	1520
1381	Qy	CACCTCTGCACCTGGAAAGGAACGTGTGTGTCAAGTTATGGGTTTCTCTTGGAAATGC	1440
1521	Db	TTTTTCTAAATGCTGGGAAGAGGTGGTGTGGCCCAATTACCAAGTGTCTCCCTTTGAGAA	1580
1441	Qy	AGCTTTGGACAGCTGCTTAACTTAGGAAGGAGCTTAATCAGTGACACAGTATCTCTGT	1500
1581	Db	AGTTTTAGAGCAGTTGCTTTAATCTAGGAAGGCACTAGTCACTGATCATGTAGACCTTGT	1640
1501	Qy	CCTGTGCCAAAAAGTTTATACATCCATCTTTTGAAGCAGTTTCTCAATPATGCATGTTAT	1560

Db 1641 CCTGTGCCAAAAGTAATCCACCACCTTTTGAGACAGTTCTTTCAGTGGCCGCCAGCTTAT 1700

QY 1561 TGCATATAGACAGAAATTTGGAGTGAGCTCTGATGGAACCCCTGACTAAATGACAGGACACACA 1620

Db 1701 GGCCATCGACAGAGTTGGGGTGACTCTGATGGAATCTCTAAGCAAAAGTCACAGGACCAAC 1760

QY 1621 GCGTATTGGATCCCTAGGCTCAATATGCTCTTAATAGTTATGAGAGTGCAAGAGATGTGG 1680

Db 1761 GCGTATTGGTTCTCTAAACCCCAATAGTTCTACTTATGAGAGTGCAAGAGATGTGG 1820

QY 1681 CACTGCAAAATTTGGCTCCCAACATTTTTTTCATCTTATCTTATGAGCAACAACTCTG 1740

Db 1821 CTCTGCAAGATTGGCTCCAAACATTTTTTTCATCTTCTCTTATGAGCCACTGTCTG 1880

QY 1741 CAGTTGCTCTCTGCAACAGAAATGACACTGCTGCGGATGAGCTGAAGCTCAGCTGTCA 1800

Db 1881 CACCTTGTCTCTCTGAGCAGAAATGACACCGCTGGGAGAGCTCAAGCTCACATGTCA 1940

QY 1801 GAGGCACTGCATGCTCTCGAGTTACACTCAAGGAACCATGGCTTTGTTGGGAGGTGG 1860

Db 1941 AACAGCAATGCGAGCTCTTGCAAGTTAACTCAAGGAACCGTGGGTTTATTTGGGAGGTGG 2000

QY 1861 CTGTACTGAACCTCAATTTGGCTGCATATATCAGACACAAAGACTCAACAGCCCAAGAAAG 1920

Db 2001 CTGTACAGAAACACACTTGGCTGCATATGTCAGACACAAAGTTTCATCAGAGCGAGAAGC 2060

QY 1921 CATTCTCAAAGATGATGAATGTAATCAAAACAGAACTTTCAATTAATGCTGAAGCAATTTTG 1980

Db 2061 TATTGTACAGATGATGGGTGTACTCAGGCAAAAGCTGCAATGTTGCTGCTGAAGCAATTTTG 2120

QY 1981 CAGTGCCTTAGAATCTGTTGGCTCTTTAGAACATGATGAGGTGAATTTCTCACTCA 2040

Db 2121 CAGTGTCTCGGAGTCCGTTGCTGGCTCTTTTGGAAACATGATGGTGAATCTCATTTGA 2180

QY 2041 CATGAAGTATGACACCTTTGGTTCAGTTTCAGGCGAGATTTCTCCTGTGTGCTAACTGGCC 2100

Db 2181 CACGAATATGACACCTTTGGTTCCTGTCACGAGATTTCTGCTCTGTGTTGTAATCTGTC 2240

QY 2101 AGATTGCTTTACAGTGGCTGTGGATTTATACAAATGACAGGAGAACTCAACTGGTTC 2160

Db 2241 AGATAGCTGTACCGGTGTGGCTGTGGTTGTGTACAAAGCAAGAGCTCAGCTGGTTC 2300

QY 2161 TTTCTTAAGACACACAGCTGCTCCATTTGTCGACAAAGCTGCTTCCACATGAAGCTGT 2220

Db 2301 TGTCTTAAGAGTACTTATCATCTTTTGCACCAAACTGCTTCCACAGCGAGCTTT 2360

QY 2221 GGGCTCAGCAGCAACCTGACCTTGGACTGTTTGAAGTCAAGAGCTTAGTGGCTACAGGT 2280

Db 2361 GGGCTCAGCAGTAACCTGACTGTGGACTGCTTCACTGCCAAGCTGAGTGGCTTACAGT 2420

QY 2281 GGGCTGAGACAGCAATTTGATTTGGGATCTTTTCATATGTTATTTGAAGATAAAACTA 2340

Db 2421 GGGCTGAGACAGCAATTTGATTTGATTTAGATCTTTTCATATGTTATTTGAAGATAAAACTA 2480

QY 2341 AGAGATAGCAATG 2353

Db 2481 ACATAAGAGAAATG 2493

RESULT 12

HS931K24/c

LOCUS

DEFINITION

Human DNA sequence from clone Rps-931K24 on chromosome 20p12. Contains ESTs, STSs, GSSs and two CpG islands. Contains a novel pseudogene, the MKKS gene for McKusick-Kaufman syndrome protein and the 5' end of a novel gene, complete sequence.

ACCESSION

AL034430

VERSION

AL034430.4 GI:11322731

KEYWORDS

HTG; CpG island; MKKS.

SOURCE

Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1 (bases 1 to 74813)

Bates,K.

Direct Submission

Submitted (12-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

On Nov 23, 2000 this sequence version replaced gi:9795189.

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence

was generated from part of bacterial clone contigs of human

Chromosome 20, constructed by the Sanger Centre Chromosome 20

Mapping Group. Further information can be found at

http://www.sanger.ac.uk/HGP/Chr20

IMPORTANT: This sequence is not the entire insert of clone RP5-931K24. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.

The true right end of clone RP5-931K24 is at 74813 in this sequence. The true left end of clone RP11-254M13 is at 26700 in this sequence. The true right end of clone RP5-1068F16 is at 100 in this sequence. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or

sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. RP5-931K24 is from the library RPi-5 constructed by the group of Pieter de Jong. For further details see

http://www.chori.org/bacpac/home.htm

VECTOR: pCYPAC2.

Location/Qualifiers

1..74813

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/chromosome="20"

/map="p12"

/clone="RP5-931K24"

/clone_lib="RPi-5"

repeat_region

/note="AluJb repeat: matches 1..153 of consensus"

367..387

repeat_region

/note="BC200 repeat: matches 126..145 of consensus"

388..500

repeat_region

/note="FLAM C repeat: matches 12..124 of consensus"

1066..1263

repeat_region

/note="MER46C repeat: matches 103..304 of consensus"

1626..1922

repeat_region

/note="AluSp repeat: matches 1..299 of consensus"

complement(1821..2342)

misc_feature

/note="match: GSS: Em:AQ281134"

2118..2257

repeat_region

/note="MERSA repeat: matches 23..189 of consensus"

2282..2403

repeat_region

/note="MERSA repeat: matches 9..154 of consensus"

complement(2331..2443)

misc_feature

/note="match: GSS: Em:AF156717"

2788..2947

repeat_region

/note="FRAM repeat: matches -2..157 of consensus"

3484..3577

repeat_region

/note="MERS3 repeat: matches 89..189 of consensus"

3600..4195

repeat_region

/note="LTR29 repeat: matches 3..619 of consensus"

complement(4233..4699)

misc_feature

LOCUS	BD276497	1274 bp	DNA	linear	PAT 17-JUL-2000
DEFINITION	143 Human Secreted Proteins.				
ACCESSION	BD276497				
VERSION	BD276497.1 GI:33086265				
KEYWORDS	JP 2002543836-A/169.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
REFERENCE	1 (bases 1 to 1274)				
AUTHORS	Soppet,D.R., Florence,K.A., Duan,R.D., Bias,C.E., Komatsoulis,G.A., Young,P.E., Moore,P.A., Steven, Ruben,M. and Rosen,C.A.				
TITLE	143 Human Secreted Proteins				
JOURNAL	Patent: JP 2002543836-A 169 24-DEC-2002; Human Genome Sciences Inc				
COMMENT	OS Homo sapiens PN JP 2002543836-A/169 PD 24-DEC-2002 PF 11-MAY-2000 JP 2000618448 PR 13-MAY-1999 US 60/134068 PI daniel r soppet,kimberly a florence,roxanne d duan,charles e pi bias, PI george a komatsoulis,paul e young,paul a moore,steven pi ruben,craig a rosen CC				
FEATURES	Key Location/Qualifiers.				
source	Location/Qualifiers 1..1274 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606"				
ORIGIN	Query Match 50.7%; Score 1253.8; DB 6; Length 1274; Best Local Similarity 99.8%; Pred. No.2:1e-209; Matches 1266; Conservative 0; Mismatches 2; Indels 1; Gaps 1;				
QY	1156	TTTGATCTCTGAGAGCCCTTTTGGCTTACAAATTCAGAAATGCTCGAAGGCCACATCATTTT	1215		
DB	1	TTTGATCTCTGAGAGCCCTTTTGGCTTACAAATTCAGAAATGCTCGAAGGCCACATCATTTT	60		
QY	1216	AGGAAGAGTTTAAATGTACCTTTAAAGGTCAAAAGGTATAGATTCCACTGTATTACC	1275		
DB	61	AGGAAGAGTTTAAATGTACCTTTAAAGGTCAAAAGGTATAGATTCCACTGTATTACC	120		
QY	1276	TGGGATCTCAATGAAATGTCAGAAAGTTCAAATTAATGAGGCTATTACCTATCAAAAAATC	1335		
DB	121	TGGGATCTCAATGAAATGTCAGAAAGTTCAAATTAATGAGGCTATTACCTATCAAAAAATC	180		
QY	1336	AACTGCCCTCAAGGTGGCACTCTTTTGTACAACCTTTTATCCGGAGACACTTCTGACACTGG	1395		
DB	181	AACTGCCCTCAAGGTGGCACTCTTTTGTACAACCTTTTATCCGGAGACACTTCTGACACTGG	240		
QY	1396	AGAAGGAACTGTGGTGGTTCAGTTATGGGGTTTCTCTGAAAGTGCAGTCTTGGACCAAGCT	1455		
DB	241	AGAAGGAACTGTGGTGGTTCAGTTATGGGGTTTCTCTGAAAGTGCAGTCTTGGACCAAGCT	300		
QY	1456	GCTTAACTCTAGGAAGGCAGCTTAATCAGTGAACCACTAGATCTTGTCTGTGCCAAAAAGT	1515		
DB	301	GCTTAACTCTAGGAAGGCAGCTTAATCAGTGAACCACTAGATCTTGTCTGTGCCAAAAAGT	360		
QY	1516	TATACATCATCTTTTGAAGCAGTTTCTCAATATGCATTCGATTATTTGCCATAGACAGAAAT	1575		
DB	361	TATACATCATCTTTTGAAGCAGTTTCTCAATATGCATTCGATTATTTGCCATAGACAGAAAT	420		
QY	1576	TGGAGTGACTCTGATGGAAACCCCTCAGCTAAAATGACAGAAACACAGCCCTATTGGATCCCT	1635		
DB	421	TGGAGTGACTCTGATGGAAACCCCTCAGCTAAAATGACAGAAACACAGCCCTATTGGATCCCT	480		
QY	1636	AGGCTCAATATGTCCTTAATAGTTATGGAAAGTGTGAAAGATGTGTGCACTGCCAAAAATTTGG	1695		
DB	481	AGGCTCAATATGTCCTTAATAGTTATGGAAAGTGTGAAAGATGTGTGCACTGCCAAAAATTTGG	540		

47002	GCAATAAATAGAAAAATGTGGTTTACAAAATAAAAACGGTCTTCACTAGTATTACCACTGA	46943
623	AGTAAGATGTCCTCGTTTGGAGCTAAGAGCCATCATTTGTGTAAGAGTGAACCACTGACA	682
46942	AGTAAGATGTCCTCGTTTGGAGCTAAGAGCCATCATTTGTGTAAGAGTGAACCACTGACA	46883
683	ACTGAGAGAGTCAGGACACACTTTCTGTCTTGAAAGAAATTTGAACATCATGCTATGGC	742
46882	ACTGAGAGAGTCAGGACCACTTTCTGTCTTGAAAGAAATTTGAACATCATGCTATGGC	46823
743	CCCTCAGGTAGGCTGAAGCAGCTGCACAATGGCTTTTGGAGGTTACGTGTGTACAACCTCA	802
46822	CCCTCAGGTAGGCTGAAGCAGCTGCACAATGGCTTTTGGAGGTTACGTGTGTACAACCTCA	46763
803	CAGTCTCAGCTGCTCAGTCACTCTTTTGGTCACACATCCCATTTTAAAGATCCTGACA	862
46762	CAGTCTCAGCTGCTCAGTCACTCTTTTGGTCACACATCCCATTTTAAAGATCCTGACA	46703
863	GCCTCCATACAGAAATCATGTCTCAAGCTTCAGTGATTTGGCTTATTTCACAGCTATTCTT	922
46702	GCCTCCATACAGNAATCATGTCTCAAGCTTCAGTGATTTGGCTTATTTCACAGCTATTCTT	46643
923	TGCTGCAACTGATTGAAAAATGTTTCAGAGATTAGGCTTCACACCCACCACTGTCAATTAGA	982
46642	TGCTGCAACTGATTGAAAAATGTTTCAGAGATTAGGCTTCACACCCACCACTGTCAATTAGA	46583
983	TTAAATAAACATCTTTTGAGTCTTTGTCATCAGTTATCTCAAGTCTGATACCTCTGTTGT	1042
46582	TTAAATAAACATCTTTTGAGTCTTTGTCATCAGTTATCTCAAGTCTGAGACCTGTGTTGT	46523
1043	CGAATCCCACTGAGTACTCAGATCCTCTCTTTTGGTGCGTAGTATATTA	1102
46522	CGAATCCCACTGAGTACTCAGATCCTCTCTTTTGGTGCGTAGTATATTA	46463
1103	ACAAAGTAAACCTGCTGTATGCTCACAGAAAGGAAACAGAGCATGTCAGTCTTTGATC	1162
46462	ACAAAGTAAACCTGCTGTATGCTCACAGAAAGGAAACAGAGCATGTCAGTCTTTGATC	46403
1163	CTGAGAGCCTTTTGTCTTACAATTCAGAAATGCTCGAAGGCCACATCATTTTAGGAAAG	1222
46402	CTGAGAGCCTTTTGTCTTACAATTCAGAAATGCTCGAAGGCCACATCATTTTAGGAAAG	46343
1223	AGTTTAATGTACCTTTTAAAGGTCAAGAGTATAGATTCCACTGATTAACCTGGGATA	1282
46342	AGTTTAATGTACCTTTTAAAGGTCAAGAGTATAGATTCCACTGTATTACCTGGGATA	46283
1283	CTCATTGAAATGTCAGAAAGTTCAAATTAATGAGGCTATTACCTATCAAAAAATCAACTGCC	1342
46282	CTCATTGAAATGTCAGAAAGTTCAAATTAATGAGGCTATTACCTATCAAAAAATCAACTGCC	46223
1343	CTCAAGGTGGCACTCTTTTGTACAACTTTATCCGGAGACACTTCTGTGACACTGGAGAAGGA	1402
46222	CTCAAGGTGGCACTCTTTTGTACAACTTTATCCGGAGACACTTCTGTGACACTGGAGAAGGA	46163
1403	ACTGTGGTGCAGTTATGGGGTTTCTCTTGAATGTCAGTCTTGGACCAGCTGCTTAAC	1462
46162	ACTGTGGTGCAGTTATGGGGTTTCTCTTGAATGTCAGTCTTGGACCAGCTGCTTAAC	46103
1463	CTAGGAAGGCAGCTAATCAGTGACCACTAGATCTTGTCTGTGCCAAAAAGTTATACAT	1522
46102	CTAGGAAGGCAGCTAATCAGTGACCACTAGATCTTGTCTGTGCCAAAAAGTTATACAT	46043
1523	CCATCTTTGAAGCAGTTTCTCAATATGCATTCGATTATTGGCCATAGACAGAAATGGAGTG	1582
46042	CCATCTTTGAAGCAGTTTCTCAATATGCATTCGATTATTGGCCATAGACAGAAATGGAGTG	45983
1583	ACTCTGATGGAAACCCCTGACTAAAATGACAGGACACAGC	1622
45982	ACTCTGATGGAAACCCCTGACTAAAATGACAGGTAAAATC	45943

RESULT 13
BD276497

may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

```

----- Project Information
Center project name: GRIS
Center clone name: CH230-20P4
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 223315 bases at least Q40
Consensus quality: 228974 bases at least Q30
Consensus quality: 232210 bases at least Q20
Estimated insert size: 226685; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
      (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*

```

ORIGIN

Query Match 36.7%; Score 908.2; DB 2; Length 259468;
Best Local Similarity 78.8%; Pred. No. 2.2e-149;

Rattus.

1 (bases 1 to 259458)

Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalobechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,K., Burch,P., Burrell,K., Calderon,B., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., DeVelam,L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpachy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorenshewa,L., Loulseg,H., Lozada,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindarte,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhinney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Mmlosavile,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Mundasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwakoilemeh,O., Okwuonu,G., Olarnpungagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L., Puozzo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Rodley,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sison,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.

TITLE
JOURNAL
REFERENCE
AUTHORS

TITLE
Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

REFERENCE
3 (bases 1 to 259468)

AUTHORS Rat Genome Sequencing Consortium.
TITLE Direct Submission
JOURNAL Submitted (09-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza Houston, TX 77030, USA

COMMENT:
On Nov 9, 2002 this sequence version replaced gi:22772218.
The sequence in this assembly is a combination of BAC reads
and whole genome shotgun sequencing reads assembled using Atlas
(<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence

Qy	259	AAAAATACCAAGTTGATAGAAAGAACTGGCTCTTTGTAGCTGGGATATCTTTTCATAGGT	318
Db	116122	AAAAACCCAGTTGGATAGAAAGAACTGGCTTCTTTGTGGCTGGGCTGACTTTCGTAGGC	116181
Qy	319	GTCCATCTTTGAAACATACCTTTTTCGACAGAGTCTGCAAGACAGTCTGTAAATTTTCAGTCT	378
Db	116182	GTCCATCTTTGAAACGTAATTAACAGAGAGCTCGCAGAGAGTCTGTGAGGTCTCAGTCC	116241
Qy	379	CAAAAGCAAAACAAAGAGATTGAAAGAGTGAAGTAAATAAATAATTTGGAATTTACTAAATTT	438
Db	116242	AGAGCAAAACAAAGGAATTTGAAGAGTGAAGCAAAATAATCTTGGAAATTTAA--T	116299
Qy	439	GTCAATTAATCAATCTATGCTGATTTAGCTTTCATAAACATTTGAACTTTTGTATTTATAGC	498
Db	116300	GTCTTTAAATTCATCTATGCTGGTTATTGTCTGAAA-----ACCACTTGATTTTGTAGCT	116354
Qy	499	CACAACTGTCGATATTTCATACCTTTTAATTTCTAAAGAAATAATTTTAAATGTTTAAACGTGA	558
Db	116355	GGTAATGCTGATGTACAGACCTTCATACCCTGGAGAAATAGTCTTTTAATGTTTCAAAATCAT	116414
Qy	559	TAATGCAATAAATAGAAAAATGTGGTTTACAAATTAATAAAACGGTCTTCACATGTTTACCAC	618
Db	116415	GACATTAATAATAATGAACAAACATGTGAAACAACTAAAGGTGTCTTCACAACTTA-CAC	116473
Qy	619	CTGAAGTAAAGATGTCTCGTTTGGAGCTAAGAAGCCATCATTTGTGTAAAGAGTGAACCACT	678
Db	116474	CTGAAGCAGGATGTCTCGTTTGTAGAGCTAAGAAGCCATCATTTGTGTAAACCTGAACCACT	116533
Qy	679	GACAACTGAGAGAGTCAAGGACCAACACTTTTCTGTCTTGAAAGAAATTTGAAACATCATGTCTA	738
Db	116534	GACAAGTGAAGAGTCAAGGTCTACACTTTTCTGTCTTGAAAGGAATTCATAGCCTCATGTCTA	116593
Qy	739	TGGCCCTCAGGTAGGCTGAAGCAGCTGCACAAATGGCTTTTGAGGTACGTGTGTGTACAAAC	798
Db	116594	TGGCCCTTCAGGGAGCTGAAGCAGCTGCACAAATGGATTAGAGGTGTGTGTGTGTACAAAC	116653
Qy	799	CTCACAGTCTCAGCTCTGCTCAGTCACTTTTGGTTCACATCCCATTTTAAAGATCCT	858
Db	116654	TTCAAGTCTCTCAGCCCTGCTTCGAAACCTTTTGGTCAACCATCCCATATTAAAGGTCTCT	116713
Qy	859	GACAGCTCCATACAGAAATCATGTGTCAAGCTTTAGTGATTTGTGGCTTATTCAAGCTAT	918
Db	116714	AACATCATCTGTGCAGAAATACGTGTGCTGCTTCACTGAGTCTGCGGCTTATTTCACAGCCAT	116773
Qy	919	TCTTTGCTGCAACCTGATTGAAATATGTTTCAGAGATTAGGCTTGACACCCACACACTGTCTAT	978
Db	116774	TCTGTGCTGCACTTTGATTGAAATATTCAGAGAAATAGGTTTGACACCCACAACTGTCTAT	116833
Qy	979	TAGATTAATAAACAATCTTTTGAAGTCTTTTGCATCAGTTTATCTCAAGTCTGATACCTGTGG	1038
Db	116834	TAAATTAATTAATACCTCTTTGGGTCTCAGTATCAGTTTATCTCAAGTCTGAAGCCGTAG	116893
Qy	1039	TTGTGCAATCCAGTGGACTTTTAGTAGTACTCAGATCCCTTCTTTGTTTGGTGGCTAGTAT	1098
Db	116894	TTGTGCAATCCAGTTGACTTCAGAAGTACACATACCTTCTCTCACTTGGTACACAGAT	116953
Qy	1099	ATTAACAAGTAAACCTGCTGTATGCTCACAGAAAGGAAACAGAGCATGTCAAGTCTTT	1158
Db	116954	CTTAACCAAGCAACAGCCTGCATGCTTACCAGAAAGGAATATAGATCATATAGGTGCTTT	117013
Qy	1159	GATCCTGAGAGCCTTTTGTCTTACAATTTCCAGAAATGCTGAAGGCCACATCATTTTAGG	1218
Db	117014	GATTTTGAAGCTTTTGTCTTACAAATTCAGAAAGTGCAGAAAGCGAATGGTTTTAGG	117073
Qy	1219	AAAGAGTTTAATGTACTTTTAAAGAGTCAAGAGTTATAGATTCACATGTATTAACCTGG	1278
Db	117074	GAAGAGTATAATTTGTTCCCTTTAAAGGGCCAAACAGTAACAGATTCCTCTGATTTACCTGG	117133
Qy	1279	GATACTCAATTCAAATGTCAGAGTTCAATTAATGAGGCTTATACCTATTCMAAAATCAAC	1338
Db	117134	ACTTCTTAATGAAGCATCAGAGTTCAATTAAGGAGATTAATACCCACCCAGAGTCAAG	117193
Qy	1339	TGCCCTCAAGGTGGCACTCTTTTGTACAACTTTTCCGGAGACACTTCTGACATCTGGAGA	1398

Db	117194	TACCCCTTAGGCTGGCAGCTCTTCTGTGCACTCTTATCTCGAGATTTTCTTAATGCTGGAGA	117253
Qy	1399	AGGAACCTGTGTGTGCTAGTATTGGGGTTTCTCTTGAAATGCAAGCTCTTGACACCAAGCTGCT	1458
Db	117254	AGGAACCTTGTGTGTGCTCAATACCAAGTGTCCCTTGAGAAATGCAGTTTTAGAGCAGTTGCT	117313
Qy	1459	TAACTTAGGAAGGCAGCTAAATCAGTGACCACTAGACTTTGTCTGTGCGCAAAAAGTTAT	1518
Db	117314	TAACTTAGGAAGACAGCTAGTCAGTGACCAATGATAGACTTTGTCTGTGCGCAAAAAGTCAT	117373
Qy	1519	ACATCCATCTTTGAAGCAGTTTCTCAATATGCATCGTATTATTGGCATAGACAGAAATTGG	1578
Db	117374	ACACCCATCTTTTGAACACAGTTCCCTCAGTGAACATCAAAATTAATGTCATCGACAGAGTTGG	117433
Qy	1579	AGTGACTCTGATGGAAACCCCTGACTTAAATGACAGGAACACAG	1621
Db	117434	AGTGACTCTGATGGAAACCTCTGAGCAAAAGTGACAGTAAAAAG	117476

Search completed: September 6, 2005, 14:56:42
Job time : 10594 secs